

UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI

Ronaldo Alves Marques Júnior

**Estudo de critérios de distribuição de vacinas
com doses limitadas - uma abordagem baseada
em redes complexas**

São João del-Rei

2021

UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI

Ronaldo Alves Marques Júnior

Estudo de critérios de distribuição de vacinas com doses limitadas - uma abordagem baseada em redes complexas

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São João del-Rei como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

Orientador: Carolina Ribeiro Xavier

Universidade Federal de São João del-Rei – UFSJ

Mestrado em Ciência da Computação

São João del-Rei

2021

Ficha catalográfica elaborada pela Divisão de Biblioteca (DIBIB)
e Núcleo de Tecnologia da Informação (NTINF) da UFSJ,
com os dados fornecidos pelo(a) autor(a)

M357e Marques Júnior, Ronaldo Alves.
Estudo de critérios de distribuição de vacinas com doses limitadas - uma abordagem baseada em redes complexas / Ronaldo Alves Marques Júnior ; orientadora Carolina Ribeiro Xavier. -- São João del Rei, 2021.
127 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Ciência da Computação) -- Universidade Federal de São João del-Rei, 2021.

1. Redes Complexas. 2. Imunização. 3. Modelos epidemiológicos. 4. Algoritmos Genéticos. I. Xavier, Carolina Ribeiro, orient. II. Título.

Ronaldo Alves Marques Júnior

Estudo de critérios de distribuição de vacinas com doses limitadas - uma abordagem baseada em redes complexas

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São João del-Rei como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

Trabalho aprovado. São João del-Rei, 5 de março de 2021:



Dra. Carolina Ribeiro Xavier
Orientador

Dr. Vinícius da Fonseca Vieira
Convidado 1

Dr. Rafael Sachetto Oliveira
Convidado 2

Dr. Rodrigo Weber dos Santos
Convidado 3

São João del-Rei
2021

*Dedico este trabalho aos meus
filhos Murilo e Henrique, fontes de
amor e inspiração únicas.
Em especial, à minha mãe “Lalesa”
(in memoriam), a fonte inesgotável
de apoio, carinho e amor ...*

Agradecimentos

Gostaria de agradecer à toda minha família e amigos que dividiram suas experiências e que me aceitaram em seu convívio. Ao meu pai Ronaldo, com suas palavras assertivas e diretas no decorrer dos meus passos. A minha irmã Renata pela companhia e sorrisos no dia a dia.

À minha amada mãe Lesa, aquela que sempre me motivou e que sempre me fez acreditar que eu podia. Quem estudava para me ensinar, que me auxiliava mesmo sem ter condições e que festejou cada passo da minha vida. Sou fruto de sua dedicação, carinho e amor incondicionais. Saudades eternas!

Aos colegas do Instituto Federal do Sudeste de Minas Gerais - Campus Barbacena, em especial aos lotados da Coordenação de Gestão de Tecnologia da Informação, que me auxiliaram com todo o possível nas atividades diárias para que eu conseguisse realizar todos os créditos necessários no curso.

Aos professores e colegas da Universidade Federal de São João del-Rei, que tão bem me receberam e me auxiliaram nessa caminhada. Aos professores Daniel Madeira, Daniel Guidoni, Fernanda Sumika, Flávio Schiavoni e Vinícius Vieira por todo empenho e dedicação à arte de ensinar. Aos professores integrantes da excelente banca examinadora pela disponibilidade em avaliar, sugerir e analisar o conteúdo deste trabalho.

Agradeço a minha orientadora Carolina Xavier, por tamanha disponibilidade, incentivo e auxílio. Sua dedicação, empenho e conhecimento foram essenciais para que eu conseguisse cumprir diversos requisitos do programa. Meu muito obrigado!

À minha esposa Aline, aquela que aceitou dividir sua vida comigo e que esteve ao meu lado durante todo este período. Obrigado por tanta resiliência, paciência, amor e dedicação à nossa família.

Por fim, aos meus filhos Murilo e Henrique, dois rapazes que desafiam os meus dias e que são a principal motivação da minha vida. Eles que me ajudaram testando códigos, lendo textos e que com palavras de que “tudo daria certo” não deixaram que eu desistisse da caminhada.

Resumo

O interesse pela dinâmica e a compreensão das características de disseminação de doenças infectocontagiosas datam de mais de dois séculos. Diversas perspectivas podem ser consideradas nesses estudos, que possuem um apreço especial pelo estudo da imunização. A escolha assertiva de grupos de indivíduos a serem imunizados pode impactar diretamente na dinâmica de contaminação, tanto na proteção de uma parte frágil da população quanto na minimização da velocidade de contágio. Tal tarefa é importante e desafiadora, especialmente em cenários com número reduzido de doses disponíveis, por envolver inúmeras combinações possíveis para tomada de decisão. O objetivo deste trabalho é propor uma metodologia para a escolha de indivíduos a serem imunizados utilizando uma abordagem em rede. Para isso, é proposta uma série de alterações e estudos aprofundados em uma metodologia baseada em algoritmo genético, tendo como linha de base um conjunto de medidas de centralidades em redes para definição dos indivíduos a serem vacinados. A efetividade de cada uma das propostas será verificada através de um estudo comparativo entre as medidas e com uma escolha aleatória de indivíduos. Os resultados obtidos por experimentos que simulam epidemias com a vacinação de acordo com cada um dos critérios serão avaliados sob as perspectivas que consideram o número de indivíduos afetados, o maior número de infectados em um mesmo instante e a taxa de contaminação estimada. Os resultados obtidos indicam que a escolha sistemática de indivíduos é uma decisão acertada para este tipo de problema e que o desempenho da metodologia baseada em algoritmos genéticos ante a algumas medidas de centralidade são equivalentes estatisticamente.

Palavras-chaves: Redes complexas, imunização, modelos epidemiológicos, algoritmos genéticos.

Abstract

The interest in the dynamics and the understanding of the spread characteristics of infectious diseases date back to more than two centuries. Several perspectives can be considered in these researches, which have a special appreciation for the study of immunization. The assertive choice of groups of individuals to be immunized can directly impact the contamination dynamics, protecting a fragile part of the population as well as in minimizing the speed of contagion. Such a task is important and challenging, especially in scenarios with a reduced number of doses available, as it involves countless possible combinations for decision making. This research aims to propose a methodology for choosing individuals to be immunized using a network approach. For that, a few changes and in-depth studies are proposed in a methodology based on a genetic algorithm, having network centrality measures as a baseline to define the individuals to be vaccinated. The effectiveness of each proposal will be verified through a comparative study between the measures and with a random choice of individuals. The results obtained by experiments that simulate epidemics with vaccination according to each of the criteria will be evaluated from the perspectives that consider the number of affected individuals, the largest number of infected individuals at the same time and the estimated contamination rate. The results indicate that the systematic choice of individuals is a correct decision for this type of problem and that the performance of the methodology based on genetic algorithms in relation to some measures of centrality are statistically equivalent.

Key-words: Complex networks, immunization, epidemiological models, genetic algorithms

Lista de ilustrações

Figura 1 – Representação dos graus da rede <i>moreno health</i>	30
Figura 2 – Diagrama de representação de alguns modelos epidemiológicos	35
Figura 3 – Comportamento epidêmico do modelo SIR aplicado em uma rede complexa.	39
Figura 4 – Diagrama de representação de um algoritmo genético	41
Figura 5 – Cruzamento de dois pontos em cromossomos com 5 genes.	44
Figura 6 – Mutação de um gene definido pela probabilidade de mutação	44
Figura 7 – Representação gráfica do AG.	48
Figura 8 – Epidemia completa. Vértices pretos foram infectados; cinzas recuperados.	51
Figura 9 – Epidemia com fim precoce. Vértices pretos foram infectados; brancos suscetíveis.	52
Figura 10 – Simulações do modelo SIR com iteração assíncrona e síncrona	53
Figura 11 – Taxa de sucesso dos experimentos de contenção direta estendidos	58
Figura 12 – Frequência dos indivíduos selecionados na análise vertical	60
Figura 13 – Intersecção de grupos de 15%, 20%, 25% e 30% de vacinados na rede <i>moreno highschool</i>	61
Figura 14 – Intersecção de grupos de 25%, 30%, 35% e 40% de vacinados na rede <i>moreno highschool</i>	61
Figura 15 – Intersecção de grupos de 15%, 20%, 25% e 30% de vacinados na rede <i>moreno innovation</i>	62
Figura 16 – Intersecção de grupos de 25%, 30%, 35% e 40% de vacinados na rede <i>moreno innovation</i>	62
Figura 17 – Histograma final do GASIR para a vacinação de 40% da população na rede <i>moreno highschool</i>	62
Figura 18 – Distribuição de grau complementar acumulada das redes estudadas	65
Figura 19 – Efeito da curva de infectados com a variação do β	70
Figura 20 – Parâmetro β inferido em uma simulação na rede <i>moreno highschool</i> com vacinação de 15% dos indivíduos	70
Figura 21 – Similaridade e intersecção dos critérios de vacinação - Redes reais x artificiais	73
Figura 22 – Percentual da população que se manteve suscetível nas redes artificiais	74
Figura 23 – Boxplots com resultados das três variáveis dependentes com vacinação de 40% na rede <i>ba400-m3</i>	75
Figura 24 – Boxplots com resultados das três variáveis dependentes com vacinação de 40% na rede <i>ba400-m6</i>	75
Figura 25 – Percentual da população que se manteve suscetível nas redes reais	77

Figura 26 – Boxplots com resultados das três variáveis dependentes com vacinação de 15% na rede <i>oz</i>	77
Figura 27 – Boxplots com resultados das três variáveis dependentes com vacinação de 40% na rede <i>innovation</i>	78
Figura 28 – Similaridade e intersecção dos critérios de vacinação entre 15% e 25% na rede <i>highschool</i>	106
Figura 29 – Similaridade e intersecção dos critérios de vacinação entre 30% e 40% na rede <i>highschool</i>	106
Figura 30 – Similaridade e intersecção dos critérios de vacinação entre 15% e 25% na rede <i>innovation</i>	107
Figura 31 – Similaridade e intersecção dos critérios de vacinação entre 30% e 40% na rede <i>innovation</i>	107
Figura 32 – Similaridade e intersecção dos critérios de vacinação entre 15% e 25% na rede <i>oz</i>	107
Figura 33 – Similaridade e intersecção dos critérios de vacinação entre 30% e 40% na rede <i>oz</i>	108
Figura 34 – Similaridade e intersecção dos critérios de vacinação entre 15% e 25% na rede <i>health</i>	108
Figura 35 – Similaridade e intersecção dos critérios de vacinação entre 30% e 40% na rede <i>health</i>	108
Figura 36 – Similaridade e intersecção dos critérios de vacinação entre 15% e 25% na rede <i>ba400-m3</i>	109
Figura 37 – Similaridade e intersecção dos critérios de vacinação entre 30% e 40% na rede <i>ba400-m3</i>	109
Figura 38 – Similaridade e intersecção dos critérios de vacinação entre 15% e 25% na rede <i>ba400-m4</i>	109
Figura 39 – Similaridade e intersecção dos critérios de vacinação entre 30% e 40% na rede <i>ba400-m4</i>	110
Figura 40 – Similaridade e intersecção dos critérios de vacinação entre 15% e 25% na rede <i>ba400-m5</i>	110
Figura 41 – Similaridade e intersecção dos critérios de vacinação entre 30% e 40% na rede <i>ba400-m5</i>	110
Figura 42 – Similaridade e intersecção dos critérios de vacinação entre 15% e 25% na rede <i>ba400-m6</i>	111
Figura 43 – Similaridade e intersecção dos critérios de vacinação entre 30% e 40% na rede <i>ba400-m6</i>	111
Figura 44 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 15% e 25% na rede <i>highschool</i>	112

Figura 45 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 15% e 25% na rede <i>highschool</i>	112
Figura 46 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 15% e 25% na rede <i>highschool</i>	113
Figura 47 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 30% e 40% na rede <i>highschool</i>	113
Figura 48 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 30% e 40% na rede <i>highschool</i>	113
Figura 49 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 30% e 40% na rede <i>highschool</i>	114
Figura 50 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 15% e 25% na rede <i>innovation</i>	114
Figura 51 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 15% e 25% na rede <i>innovation</i>	114
Figura 52 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 15% e 25% na rede <i>innovation</i>	115
Figura 53 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 30% e 40% na rede <i>innovation</i>	115
Figura 54 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 30% e 40% na rede <i>innovation</i>	115
Figura 55 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 30% e 40% na rede <i>innovation</i>	116
Figura 56 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 15% e 25% na rede <i>oz</i>	116
Figura 57 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 15% e 25% na rede <i>oz</i>	116
Figura 58 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 15% e 25% na rede <i>oz</i>	117
Figura 59 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 30% e 40% na rede <i>oz</i>	117
Figura 60 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 30% e 40% na rede <i>oz</i>	117
Figura 61 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 30% e 40% na rede <i>oz</i>	118
Figura 62 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 15% e 25% na rede <i>health</i>	118
Figura 63 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 15% e 25% na rede <i>health</i>	118

Figura 64 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 15% e 25% na rede <i>health</i>	119
Figura 65 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 30% e 40% na rede <i>health</i>	119
Figura 66 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 30% e 40% na rede <i>health</i>	119
Figura 67 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 30% e 40% na rede <i>health</i>	120
Figura 68 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 15% e 25% na rede <i>ba400-m3</i>	120
Figura 69 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 15% e 25% na rede <i>ba400-m3</i>	120
Figura 70 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 15% e 25% na rede <i>ba400-m3</i>	121
Figura 71 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 30% e 40% na rede <i>ba400-m3</i>	121
Figura 72 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 30% e 40% na rede <i>ba400-m3</i>	121
Figura 73 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 30% e 40% na rede <i>ba400-m3</i>	122
Figura 74 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 15% e 25% na rede <i>ba400-m4</i>	122
Figura 75 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 15% e 25% na rede <i>ba400-m4</i>	122
Figura 76 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 15% e 25% na rede <i>ba400-m4</i>	123
Figura 77 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 30% e 40% na rede <i>ba400-m4</i>	123
Figura 78 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 30% e 40% na rede <i>ba400-m4</i>	123
Figura 79 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 30% e 40% na rede <i>ba400-m4</i>	124
Figura 80 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 15% e 25% na rede <i>ba400-m5</i>	124
Figura 81 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 15% e 25% na rede <i>ba400-m5</i>	124
Figura 82 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 15% e 25% na rede <i>ba400-m5</i>	125

Figura 83 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 30% e 40% na rede <i>ba400-m5</i>	125
Figura 84 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 30% e 40% na rede <i>ba400-m5</i>	125
Figura 85 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 30% e 40% na rede <i>ba400-m5</i>	126
Figura 86 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 15% e 25% na rede <i>ba400-m6</i>	126
Figura 87 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 15% e 25% na rede <i>ba400-m6</i>	126
Figura 88 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 15% e 25% na rede <i>ba400-m6</i>	127
Figura 89 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 30% e 40% na rede <i>ba400-m6</i>	127
Figura 90 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 30% e 40% na rede <i>ba400-m6</i>	127
Figura 91 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 30% e 40% na rede <i>ba400-m6</i>	128

Lista de tabelas

Tabela 1 – Nomenclatura por área de estudo	27
Tabela 2 – Dados gerais das redes utilizadas para para simulações	47
Tabela 3 – Quantidade de vacinas disponíveis por rede	57
Tabela 4 – Conjunto de parâmetros selecionados para simulação de epidemias de <i>origem desconhecida</i>	59
Tabela 5 – Parâmetros do AG aplicados para os experimentos	60
Tabela 6 – Dados gerais das redes utilizadas para simulações	66
Tabela 7 – Redes analisadas: infectados e vacinas	67
Tabela 8 – Melhores critérios avaliados pela mediana em redes artificiais	76
Tabela 9 – Melhores critérios avaliados pela mediana em redes reais.	78
Tabela 10 – Calibragem do AG após as mudanças propostas	91
Tabela 11 – Critérios que apresentaram menor mediana para a variável dependente <i>pico</i>	97
Tabela 12 – Critérios que apresentaram menor mediana para a variável dependente <i>beta</i>	100
Tabela 13 – Critérios que apresentaram menor mediana para a variável dependente <i>afetados</i>	103

Lista de Algoritmos

1	Algoritmo que representa a dinâmica das simulações de efetividade	69
---	---	----

Lista de abreviaturas e siglas

AG	Algoritmo genético
FDA	Função de distribuição acumulada
CDF	<i>Cumulative Distribution Function</i>
CCDF	<i>Complementary Cumulative Distribution Function</i>
RBO	<i>Rank-Biased Overlap</i>
SIR	modelo suscetível, infectado, recuperado;
SSAG	Algoritmo genético que utiliza “ <i>steady-state</i> ” como estratégia de reprodução

Lista de símbolos

S	Conjunto ou compartimento de indivíduos suscetíveis
I	Conjunto ou compartimento de indivíduos infectados
R	Conjunto ou compartimento de indivíduos recuperados
E	Conjunto ou compartimento de indivíduos expostos
N	Tamanho de uma população
R_0	Número básico de reprodução
β	Taxa de contágio de uma epidemia simulada
γ	Taxa de recuperação de uma epidemia simulada
μ	Taxa de infecção pós exposição de uma epidemia simulada
i, j	Índices que representam uma posição ordinal de um elemento em um conjunto
$\rho_k^S(t)$	Densidade parcial no instante t de indivíduos suscetíveis com grau k
$\rho_k^I(t)$	Densidade parcial no instante t de indivíduos infectados com grau k
$\rho_k^R(t)$	Densidade parcial no instante t de indivíduos recuperados com grau k
h_i	Taxa de infecção espontânea de um vértice i
V	Conjunto de vértices de uma rede
E	Conjunto de arestas de uma rede
A	Matriz de adjacências
$n, V $	Cardinalidade de vértices em uma rede complexa
$m, E $	Cardinalidade de arestas em uma rede complexa
k	Grau de um vértice
$\langle k \rangle, \bar{k}$	Grau médio de uma rede
ρ	Densidade da rede complexa
c	Coefficiente de agrupamento

$p(k)$	Fração dos vértices de grau k
r	Coefficiente de (dis)assortatividade
$\langle k \rangle$	grau médio de uma rede
W	Conjunto de soluções do espaço de busca um algoritmo de otimização
s_{ag}	Parâmetro que define o <i>gap</i> do modo “ <i>steady state</i> ” do AG.
g_{ag}	Parâmetro que define o número gerações utilizadas no AG.
m_{ag}	Parâmetro que define a taxa de mutação do AG.
p_{ag}	Parâmetro que define o tamanho da população do AG.
p	P-valor ou valor de significância de um teste estatístico.
pv	Percentual de vacinados aplicados a uma população.
\mathcal{C}_{pv}^D	Conjunto de indivíduos selecionados a partir de sua centralidade por grau para uma vacinação de pv da população.
\mathcal{C}_{pv}^B	Conjunto de indivíduos selecionados a partir de sua centralidade por <i>betweenness</i> para uma vacinação de pv da população.
\mathcal{C}_{pv}^P	Conjunto de indivíduos selecionados a partir de sua centralidade <i>page-rank</i> para uma vacinação de pv da população.
\mathcal{C}_{pv}^E	Conjunto de indivíduos selecionados a partir de sua centralidade <i>eigen-vector</i> para uma vacinação de pv da população.
\mathcal{C}_{pv}^G	Conjunto de indivíduos selecionados a partir da metodologia GASIR para uma vacinação de pv da população.
\mathcal{C}_{pv}^R	Conjunto de indivíduos selecionados randomicamente para uma vacinação de pv da população.
\mathcal{C}^\emptyset	Ausência de indivíduos vacinados.
X	Conjunto de indivíduos candidatos à infecção na rede
Z	Conjunto que contém agrupamentos de infectados.

Sumário

1	INTRODUÇÃO	21
1.1	Objetivos	22
1.2	Justificativa	23
1.3	Contribuições	24
1.4	Trabalhos relacionados	24
1.5	Organização do texto	25
2	REFERENCIAL TEÓRICO	27
2.1	Redes Complexas	27
2.1.1	Definições gerais	27
2.1.2	Medidas estruturais em redes complexas	28
2.1.2.1	Grau	28
2.1.2.2	Coeficiente de agrupamento	29
2.1.2.3	Distribuição de grau	29
2.1.2.4	Correlação de grau	30
2.1.3	Processos de formação de redes aleatórias	31
2.1.3.1	Redes aleatórias	31
2.1.3.2	Redes de mundo pequeno	31
2.1.3.3	Redes livres de escala	32
2.1.4	Redes homogêneas e heterogêneas	32
2.1.5	Centralidade	33
2.1.5.1	Grau	33
2.1.5.2	Betweenness	33
2.1.5.3	Autovetor	33
2.1.5.4	Pagerank	34
2.2	Epidemiologia Matemática	34
2.2.1	Modelo SIR determinístico	35
2.2.2	O modelo SIR em redes complexas	36
2.3	Algoritmos genéticos	39
2.3.1	Conceitos gerais	40
2.3.2	Representação de soluções	41
2.3.3	Função objetivo	42
2.3.4	Tratamento de restrições	42
2.3.5	Sujeitos no AG	42
2.3.6	População inicial	43

2.3.7	A seleção	43
2.3.8	Cruzamento	43
2.3.9	Mutação	44
2.3.10	Cr�terios de sobreviv�ncia ou reprodu�o	44
3	A METODOLOGIA GASIR	46
3.1	Defini�es, ferramental e o cen�rio experimental	46
3.2	Modelagem da metodologia	48
3.3	A fun�o objetivo	50
3.4	O algoritmo gen�tico	54
3.5	Metodologia experimental	55
3.6	Resultados	57
3.6.1	Experimentos que envolvem uma epidemia de origem conhecida	57
3.6.2	Experimentos que envolvem uma epidemia de origem desconhecida	59
3.7	Considera�es finais do cap�tulo	63
4	EFETIVIDADE DOS CR�TERIOS DE VACINA�O	64
4.1	A metodologia dos experimentos	64
4.1.1	Descri�o das redes avaliadas	64
4.1.2	Metodologia aplicada para os experimentos	66
4.2	Resultados	72
4.2.1	Similaridades e intersec�o	72
4.2.2	An�lise estat�stica das vari�veis dependentes	73
4.3	Considera�es finais do cap�tulo	79
5	CONCLUS�ES	81
5.1	Limita�es e trabalhos futuros	82
	REFER�NCIAS	83
	AP�NDICES	90
	AP�NDICE A – TABELAS COM RESULTADOS DE ALGUNS EXPERIMENTOS	91
	AP�NDICE B – CORRELA�O ENTRE AS HEUR�STICAS DE VACINA�O	106
	AP�NDICE C – BOXPLOT DAS VARI�VEIS DEPENDENTES CONSIDERADAS NO ESTUDO.	112

1 Introdução

Dentre diversas definições, considera-se que a epidemiologia é a ciência que estuda o adoecimento de populações e que analisa os fatores e a distribuição dessas doenças com o objetivo de propor ações de prevenção, controle ou erradicação através de indicadores que auxiliem no planejamento e na administração de ações planejadas (ROUQUAYROL; FILHO, 2003). Segundo Porta (2014), a epidemiologia estuda a frequência de uma doença na mesma área, população e estação do ano.

O controle e acompanhamento de doenças infectocontagiosas é um constante desafio que governos e agências de saúde precisam enfrentar. Sistemas de vigilância epidemiológica tornam-se importantes para identificação ágil de epidemias, deste modo medidas de controle e prevenção podem ser implementadas. A definição de estratégias que garantam a prevenção de doenças dessa natureza normalmente envolvem a utilização de mecanismos que garantam a imunização.

A imunização busca o aumento da resistência de um indivíduo a infecções. A vacinação é um dos métodos utilizados para se atingir essa condição. Segundo Feijo e Safadi (2006) e CDC (2011) a vacinação é um dos principais fatores de promoção de saúde e prevenção de doenças. O estudo realizado em Li et al. (2021) concluiu que campanhas de vacinação para controle de doenças infectocontagiosas geradas por dez patógenos diferentes evitaram 37 milhões de mortes (intervalo de confiança de 95%) entre os anos de 2000 e 2019.

Os custos envolvidos na imunização através da vacinação podem causar inúmeros empecilhos para a execução de ações rápidas e capazes de impactar diretamente na disseminação de doenças. Fatores como a escassez de recursos, dificuldades de estocagem e distanciamento geográfico podem criar um cenário árduo para a tomada de decisão.

O interesse pela dinâmica e a necessidade de se compreender as características de proliferação de doenças motivaram o surgimento da área de epidemiologia matemática. Busca-se a criação de modelos capazes de reproduzir o processo dinâmico de disseminação da doença. Dentre vários, o SIR (Suscetível-Infetado-Recuperado) se destaca devido a sua aplicação em diferentes contextos e doenças, conforme descrito em Rodrigues (2016). O modelo utiliza de três compartimentos mutuamente exclusivos que dão origem ao seu nome e que indicam o estado do indivíduo ante a doença. Assim, indivíduos infectados estão hábeis a infectar outros indivíduos e também têm a possibilidade de se recuperar da doença.

Computacionalmente o modelo SIR pode ser implementado em redes complexas de forma a registrar interações estocásticas entre indivíduos da população. Cada um dos

vértices da rede representa um indivíduo que pertence a população e as arestas representam o contato entre os indivíduos. Todos os vizinhos de um vértice que está infectado podem assumir essa condição com uma probabilidade β . Esses mesmos indivíduos infectados podem também se recuperar da doença com a probabilidade γ . Inúmeras pesquisas utilizam desses recursos que reproduzem a complexidade das relações entre indivíduos vistas em populações reais.

Todo esse ferramental pode ser utilizado para aproximar o comportamento epidêmico de diversas doenças, o que viabiliza inclusive a análise de estratégias distintas de vacinação. A seleção de indivíduos a serem vacinados é análoga à escolha de vértices influentes em uma rede complexa, problema caracterizado como NP-difícil conforme visto em [Kempe et al. \(2005\)](#). Essa seleção depende da combinação de diversos indivíduos para que se verifique qual configuração afeta ou inibe o espalhamento da epidemia em uma rede.

Problemas de otimização dessa espécie podem ser resolvidos com a utilização de algoritmos genéticos. Algoritmos genéticos (AG) atuam na busca de soluções para um problema de otimização global e que reproduzem computacionalmente critérios de evolução de espécies atuando com uma população de soluções candidatas.

Este trabalho realiza o estudo sobre diferentes critérios de vacinação que podem ser adotados em cenários controlados onde existe a escassez do número doses para imunização das populações representadas por redes complexas. Assume-se a hipótese de que é possível escolher indivíduos para imunização que, graças aos seus relacionamentos, diminuem a taxa de contágio ou o número de indivíduos afetados pela doença. Uma metodologia baseada em algoritmos genéticos para a escolha de indivíduos a serem imunizados terá sua eficácia avaliada e comparada a outros critérios de seleção de indivíduos baseados em medidas de redes complexas. Nos resultados observou-se que a utilização de algum critério sistemático de escolha de indivíduos é uma decisão acertada e que, nos cenários experimentados, a metodologia teve desempenho similar a outros critérios que possuem custo computacional menor.

1.1 Objetivos

O objetivo deste trabalho é comparar diferentes critérios de escolhas de indivíduos a serem imunizados em cenários com número restrito de doses para imunização. A análise será feita através da observação dos impactos que estas diferentes escolhas podem causar no processo dinâmico de disseminação de uma doença em uma rede complexa. O modelo epidemiológico SIR é utilizado nos experimentos que simulam uma epidemia, onde são vacinados os indivíduos escolhidos a cada critério e cuja infecção é realizada aleatoriamente.

Dentre os critérios avaliados, uma metodologia de otimização baseada em algoritmos genéticos é considerada para a proposição do grupo de indivíduos ideal. A metodologia se apoia na execução de múltiplas simulações da epidemia, com diferentes infectados, na busca de destacar os indivíduos que frequentemente surgem como solução no anseio de minimizar o número de indivíduos afetados pelas simulações.

Resumidamente, o presente trabalho tem como objetivos:

- Propor e analisar a viabilidade de uso do algoritmo de otimização populacional como forma de prover, através de diversas simulações, um conjunto de vacinados ideal capaz de reduzir os efeitos da disseminação de uma doença em uma rede complexa.
- Propor uma metodologia de experimentação e análise que auxilie na avaliação de diferentes critérios de vacinação e que considerem:
 - Analisar os níveis de similaridade e de intersecção dos grupos de indivíduos propostos para cada critério de vacinação proposto. Diferentes percentuais de vacinação serão considerados na avaliação.
 - Analisar estatisticamente os efeitos observados que os diferentes critérios de vacinação desencadearam em simulações epidêmicas, observados através de informações que considerem a taxa de contágio e o número de indivíduos afetados pela doença nos experimentos.
- Concluir sobre a efetividade e a viabilidade dos critérios de vacinação adotados, observando-se características das redes utilizadas e dos impactos gerados no processo dinâmico da disseminação da doença nestas redes.

1.2 Justificativa

A vacinação é o principal mecanismo de imunização utilizado para a promoção de saúde e controle de epidemias. Em meio à preocupação com os números de óbitos e infecções frutos da pandemia de COVID-19, o trabalho de [Oliveira et al. \(2021\)](#) discute o quanto que a taxa de imunização impacta na disseminação e controle da pandemia.

Modelos epidemiológicos clássicos consideram o princípio de “*mistura homogênea*” onde qualquer indivíduo tem a mesma probabilidade de contato com qualquer outro indivíduo da população. Esse princípio generaliza quaisquer características geográficas que são observadas em populações reais. Neste trabalho discute-se uma abordagem heterogênea, que busca representar a complexidade do contato de diferentes pessoas e as estruturas geográficas das populações através da representação de redes complexas. Modelos epidemiológicos aplicados nessas relações consideram toda a estrutura de contato entre os

indivíduos na disseminação da doença, através da aleatoriedade determinada pela probabilidade de contágio β e a probabilidade de recuperação γ . A utilização desses recursos computacionais busca aproximar a especificidade de situações reais através da modelagem e experimentação em ambientes controlados.

Tal representação permite considerar a importância de cada indivíduo na sua população e averiguar se uma escolha sistemática pode impactar na velocidade de contágio e no número de indivíduos afetados por uma doença, mesmo com um número reduzido de doses para vacinação.

1.3 Contribuições

Como contribuições deste trabalho, destaca-se:

- A aplicação e a investigação de uma metodologia de otimização que utilize de algoritmos genéticos para a escolha de indivíduos que avalie “in silico” a importância destes a partir de um método dinâmico de disseminação de doenças em redes complexas.
- Especificação de critérios para a calibragem de parâmetros e análise dos resultados obtidos pela metodologia de otimização.
- Implementação de uma rotina experimental para a aquisição de dados epidêmicos que considere diferentes grupos de vacinados para discussão estatística dos resultados obtidos.
- Análise de similaridade e intersecção de grupos de indivíduos propostos por critérios de vacinação analisados em redes complexas de diferentes características topológicas.
- Implementação de uma metodologia sistemática para avaliação de grupos de indivíduos ou vértices, a partir dos resultados de processos dinâmicos aplicados em redes complexas, utilizando-se de análise estatística das variáveis dependentes consideradas.

1.4 Trabalhos relacionados

O modelo SIR é comumente utilizado para predição e caracterização de doenças. Em meio a atual pandemia de COVID-19 provocada por um coronavírus, diversos trabalhos como [Ndiaye et al. \(2020\)](#), [Reis et al. \(2020a\)](#) e [Reis et al. \(2020b\)](#) utilizam do mesmo modelo para este fim, adicionando características que consideram óbitos e período de incubação da doença.

Diversas pesquisas utilizam modelos epidemiológicos para análise de estratégias de imunização. Como exemplo tem-se o trabalho de [Tanner et al. \(2008\)](#) que utilizou programação estocástica para encontrar políticas ótimas de vacinação e [Shulgin et al. \(1998\)](#) que a partir de modelo SIR propõe o princípio de vacinação em pulso, capaz de, segundo o estudo, erradicar doenças com baixa vacinação. O interesse é ainda mais frequente quando modelos epidemiológicos são aplicados em conjunto de elementos de redes complexas. Alguns trabalhos como [Nguyen e Carlson \(2016\)](#), [Liu et al. \(2017\)](#) e [Li et al. \(2019\)](#) avaliam e propõem diferentes estratégias de imunização, seja ela passiva ou ativa. Na imunização passiva vê-se a escolha direta dos indivíduos a serem vacinados. Já na ativa realiza-se um acompanhamento frequente da condição de contágio na rede e, após o disparo de algum critério de observação, define-se frações da população a serem vacinadas. Em [Abbassi e Heidari \(2011\)](#) utilizou-se de um algoritmo de busca local para propor a escolha de indivíduos candidatos a imunização em uma rede de contatos real com milhares de vértices e que reduziu o número de infectados. O trabalho [Pastor-Satorras et al. \(2015\)](#) estende o estudo de redes complexas e modelos epidemiológicos, em uma revisão detalhada que considera aspectos físicos e matemáticos de diversos modelos.

Os modelos epidemiológicos podem ser ainda adaptados a outros processos dinâmicos em redes, como o espalhamento de outras característica. O trabalho de [Woo et al. \(2011\)](#) utilizou o SIR para análise de assuntos violentos em fóruns e [Kandhway e Kuri \(2014\)](#) utilizou princípio equivalente para modelagem de campanhas eleitorais ou publicitárias.

1.5 Organização do texto

A estrutura textual deste trabalho foi organizada conforme descrito a seguir.

O Capítulo 2 apresenta todo o referencial teórico utilizado como ferramental para o desenvolvimento desta pesquisa, organizado em tópicos que tratam de redes complexas, epidemiologia matemática e algoritmos genéticos.

O Capítulo 3 apresenta a metodologia de otimização e todas as suas especificidades. Uma série de tópicos abrange os desafios e as diretrizes da metodologia. Experimentos descrevem os critérios para a escolha dos parâmetros do algoritmo e o nível de estabilidade da metodologia proposta.

O Capítulo 4 discute diferentes critérios de seleção de indivíduos para vacinação. Uma primeira análise discute o nível de similaridade dos grupos selecionados pelos critérios. Essas impressões são estendidas para a análise do impacto que a vacinação desses indivíduos causa na disseminação da doença. Uma série de simulações epidêmicas provê amostras que consideram três diferentes variáveis dependentes. A análise estatística dessas amostras auxilia na interpretação dos critérios que impactam na disseminação da

epidemia positivamente, seja diminuindo a taxa de contágio ou minimizando o número de indivíduos infectados pela doença.

O Capítulo 5 resume os resultados obtidos pelas duas vertentes do trabalho, apresenta algumas de suas limitações e a proposta de trabalhos futuros.

No Apêndice A vê-se a tabela com os dados obtidos no processo de calibragem do AG tratados no Capítulo 3. As amostras obtidas nos experimentos realizados e que consideram cada um dos critérios de vacinação estudados no Capítulo 4 são fonte de diversas análises deste trabalho. No Apêndice A, para cada variável dependente analisada, tabelas divulgam os critérios com a menor mediana agrupados por percentual de vacinação e rede. Exibem ainda as medidas de dispersão, posição e informações dos testes não-paramétricos de equivalência estatística das amostras. No Apêndice B ilustra-se diversos mapas de calor que representam os níveis de similaridade e a intersecção dos indivíduos selecionados nos critérios analisados, agrupados por rede e percentual de vacinação. No Apêndice C lista-se uma série de “*boxplots*” que ilustram os dados das amostras de cada variável dependente considerada.

2 Referencial teórico

2.1 Redes Complexas

Sistemas complexos definem uma série de componentes e as suas interações entre si. O entendimento de suas características e comportamentos tornou-se uma linha de estudo bastante utilizada. Um dos desafios encontrados foi o de se representar adequadamente toda a informação que os componentes e suas interações geram. Surge então o estudo de Redes Complexas, que modela as interações entre os elementos do sistema como grafos e realiza diferentes estudos topológicos e numéricos na busca de padrões e tendências.

Esta linha de estudos teve um acentuado crescimento de interesse no início do século XX, quando se avalia o número de citações a dois trabalhos essenciais e que introduziram este estudo: [Erdos e Rényi \(1960\)](#) e [Granovetter \(1977\)](#).

A utilização deste ferramental fomentou estudos que utilizam dos conceitos de redes complexas em sistemas reais como Internet, redes neurais, finanças, redes sociais, redes de comunicação, redes de proteínas, distribuição de redes elétricas e de entregas e outras ([BARABASI, 2016](#)).

2.1.1 Definições gerais

Segundo [Barabasi \(2016\)](#) a representação em rede cria uma linguagem comum para o estudo de sistemas complexos de naturezas diferentes.

É importante salientar que na literatura científica existe alguma diferenciação dos termos utilizados conforme a área de estudo. Para garantir a correção e o entendimento destas características, na [Tabela 1](#) lista-se a associação das terminologias por área de estudo, conforme abordado nos trabalhos de [Newman \(2003\)](#), [Brito \(2012\)](#) e [Barabasi \(2016\)](#). Existem diferenças bastante sutis entre estas terminologias, mas que são representativas no contexto a que se aplicam.

Tabela 1 – Nomenclatura por área de estudo

Redes complexas	Matemática	Física
Rede	Grafo	Rede
Nó	Vértice	Sítios
Conexões	Aresta	Conexões

Formalmente, uma rede ou grafo representado por $G(V, E)$ é formado pelo con-

junto V que contém n vértices e pelo conjunto E que contém m arestas. Cada aresta registra a interação entre dois vértices e pode possuir ou não a indicação de direção. Essa característica classifica a rede como direcionada ou como não direcionada.

A existência de arestas entre os vértices pode ser matematicamente representada por uma matriz de adjacências A quadrada e de ordem $n \times n$, onde n representa o número de vértices do grafo.

$$A = \begin{cases} a_{ij} = 1 & \text{caso haja aresta entre os vértices } i \text{ e } j \\ a_{ij} = 0 & \text{caso contrário} \end{cases}$$

Dois vértices i e j que possuem interações são chamados de adjacentes ou vizinhos. Sempre que o grafo for não direcionado, teremos $a_{ij} = a_{ji}$. Se o grafo for direcionado, pode-se ter $a_{ij} \neq a_{ji}$ caso a conexão não ocorra nos dois sentidos. Portanto a matriz A é simétrica para redes não direcionadas e pode ser assimétrica para redes direcionadas.

Essas conexões podem possuir pesos, que representam as características das interações. Considerada ainda a matriz de adjacência A , os valores recebem na posição a_{ij} os pesos das interações entre os vértices i e j .

Outro conceito importante é o de componente. Um componente representa um subgrafo isolado e não vazio. Em redes complexas, o componente gigante determina o maior subgrafo conectado e que possui a maioria dos vértices da rede. Tal definição é frequentemente utilizada no estudo e na compreensão de características topológicas das redes.

2.1.2 Medidas estruturais em redes complexas

2.1.2.1 Grau

O grau é uma das principais medidas utilizadas no estudo de redes complexas. Um vértice $i \in V$ possui grau k_i que representa o número de arestas que incidem nele. Como cada aresta conecta dois vértices, espera-se que a soma de todos os graus da rede deva ser registrado pelo dobro do número de arestas existentes. Para redes direcionadas, esta medida é decomposta através da contagem de arestas de entrada k_i^{in} e de arestas de saída k_i^{out} . Decorre desta definição o interesse pelo grau médio \bar{k} ou $\langle k \rangle$ que é dado pela média aritmética do grau de todos os vértices da rede. A Equação 2.1 representa numericamente a soma de todos os graus e o grau médio de uma rede complexa.

$$\begin{aligned} \sum_{i \in V} k_i &= 2m \\ \bar{k} &= \frac{1}{n} \sum_{i \in V} k_i = \frac{2m}{n} \end{aligned} \tag{2.1}$$

onde

$$\begin{aligned} m &= \text{número total de arestas } |E| \\ n &= \text{número total de vértices } |V| \\ \bar{k} &= \text{grau médio da rede} \\ k_i &= \text{grau do vértice } i \end{aligned}$$

Medidas que consideram o número de arestas também auxiliam na caracterização das redes. A densidade ρ calcula a razão entre o número de arestas existentes e o total de arestas possíveis de uma rede. Cada um dos n vértices pode se relacionar com $n - 1$ vértices da rede. Por consequência, podem existir na rede $n(n - 1)/2$ arestas. A divisão por 2 decorre da contagem em dobro de arestas. Com isso, a densidade pode ser definida pela Equação 2.2.

$$\rho = \frac{m}{n(n - 1)/2} = \frac{\bar{k}}{n - 1} \quad (2.2)$$

2.1.2.2 Coeficiente de agrupamento

O coeficiente de agrupamento local, ou coeficiente de *clustering*, mede a redundância de arestas ao redor de um vértice. Empiricamente, calcula-se o número de arestas que os vizinhos de um vértice i possuem entre si e o máximo de arestas possíveis para este mesmo conjunto de vértices. A Equação 2.3 mostra matematicamente esta representação.

$$c_i = \frac{2|E_i|}{k_i(k_i - 1)} \quad (2.3)$$

onde

$|E_i|$ = cardinalidade do conjunto de arestas existentes entre os vizinhos de i .

O coeficiente de *clustering* da rede pode ser definido pela média aritmética dos coeficientes locais e dado pela Equação 2.4.

$$c = \frac{1}{n} \sum_{i \in V} C_i \quad (2.4)$$

Os trabalhos de Fagiolo (2007) e Barabasi (2016) discutem todas as nuances e características destas medidas em detalhes.

2.1.2.3 Distribuição de grau

A distribuição de grau busca representar o panorama de conectividade de uma rede. Tem-se então uma medida escalar $p(k)$ que busca representar uma fração dos vértices que possuem determinado grau k .

$$p(k) = \frac{n_k}{n} \quad (2.5)$$

onde

n_k = número de vértices de grau k

n = número total de vértices

Segundo Newman (2003), a distribuição de grau pode ser observada graficamente através de histogramas, embora esta possa se apresentar bastante ruidosa dado a possibilidade de existir uma diferença escalar grande entre os dados obtidos. No intuito de se facilitar a interpretação destes dados, pode-se ajustar a escala do histograma plotando-se em escala logarítmica ou utilizar-se de uma função de distribuição acumulada $P(k)$ (CDF - *Cumulative Distribution Function*) que deve ser calculada através da Equação 2.6. Diz-se que $P(k)$ representa a probabilidade de um vértice escolhido ao acaso possuir grau menor ou igual a k .

$$P(k) = \sum_{k'=k}^{\infty} p(k') \quad (2.6)$$

No estudo de redes complexas é muito comum a representação da função de distribuição acumulada complementar (CCDF - *Complementary Cumulative Distribution Function*). A diferença é a representação é dada pela probabilidade de um vértice ter grau maior que k . A Figura 1 ilustra um histograma e uma CCDF de uma rede complexa.

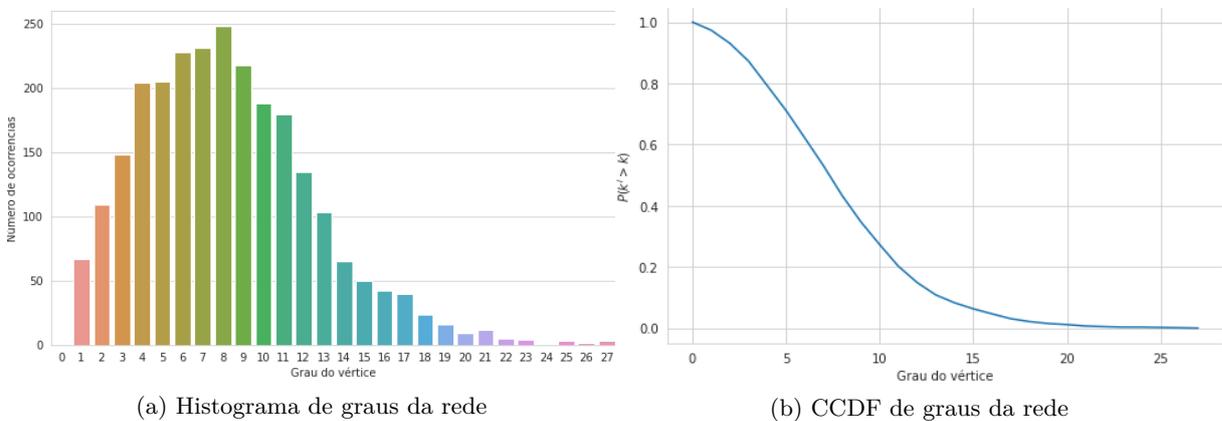


Figura 1 – Representação dos graus da rede *moreno health*

2.1.2.4 Correlação de grau

A correlação de grau busca estabelecer a tendência de conexão entre vértices segundo suas características. Assim, a tendência de conexão entre vértices de características semelhantes é definida como assortatividade. Quando a tendência é de conexão entre vértices com características opostas, define-se a disassortatividade.

Em redes complexas, a característica mais frequentemente usada para o estudo de correlação é o grau. Assim, diz-se que a rede é assortativa quando existe uma tendência

de vértices de grau alto conectarem entre si. Uma rede pode ser dita como disassortativa quando existe a tendência de vértice de grau baixo se conectarem aos de grau mais alto

O trabalho de [Newman \(2002\)](#) propôs uma medida quantitativa capaz definir o grau de disassortatividade através de uma correlação de Pearson dada pela Equação 2.7.

$$r = \frac{m^{-1} \sum_i k_i x_i - \left[m^{-1} \sum_i \frac{k_i + x_i}{2} \right]^2}{m^{-1} \sum_i \left(\frac{k_i^2 x_i^2}{2} \right) - \left[m^{-1} \sum_i \frac{k_i + x_i}{2} \right]^2} \quad (2.7)$$

onde

k_i, x_i = grau do vértices que terminam a aresta i tal que $i \in E$

m = cardinalidade do conjunto de arestas E

Assim como em uma correlação de Pearson, o coeficiente varia entre $-1 \leq r \leq 1$ onde a rede é definida como assortativa se $r > 0$ e disassortativa se $r < 0$. Quando $r = 0$ diz-se que a rede é neutra.

2.1.3 Processos de formação de redes aleatórias

Na literatura de redes complexas vê-se diversos modelos que podem ser aplicados e que são destacados nas seções a seguir.

2.1.3.1 Redes aleatórias

A formação de redes aleatórias é fruto da utilização de modelos que buscam reproduzir a complexidade sistemas reais através do posicionamento aleatório de arestas entre os vértices de uma rede. A cada passo vértices da rede podem receber uma aresta segundo uma probabilidade p ou de acordo com o um número L de arestas estipulado para a rede.

Estes diferentes critérios para a formação da rede decorrem dos trabalhos de [P. Erdos e Rényi \(1959\)](#) e de [Gilbert \(1959\)](#) e que consideram as $\frac{N(N-1)}{2}$ conexões possíveis de uma rede de N vértices.

$G(N, L)$: onde os N vértices são conectados a partir de L links posicionados aleatoriamente. ([P. Erdos; Rényi, 1959](#))

$G(N, p)$: onde os N vértices são conectados com uma probabilidade p . ([GILBERT, 1959](#))

2.1.3.2 Redes de mundo pequeno

Também conhecidas como redes *Small World* surgiram a partir das observações de [Watts e Strogatz \(1998\)](#) que percebeu que muitas redes eram altamente conectadas,

mas que os vértices possuíam poucas arestas. Assim formou-se um modelo de rede que a partir de uma rede regular, de agrupamento alto, adicionava-se aleatoriedade com a adição de arestas resultando na diminuição do caminho médio. No trabalho de Newman (2003) realizou-se uma revisão de diversas características deste modelo, abrangendo medidas como coeficiente de aglomeração, distribuição de grau e caminho médio da rede.

2.1.3.3 Redes livres de escala

Estas redes também são conhecidas como *Scale-Free* e são fruto do modelo que surgiu após as investigações realizadas em Barabasi e Albert (1999), que procuraram verificar se a *World Wide Web* apresentava características das redes *Small World*. Foi verificado que a conexão entre novos vértices não segue uma distribuição de probabilidade uniforme. Verifica-se a existência de vértices que possuem elevado grau, chamados de *hubs*, coexistindo com a vasta maioria de grau menor. A distribuição de grau $P(k)$ apresenta uma “cauda pesada” aproximando-se do comportamento de uma lei de potência, com forma $P(k) \approx k^{-\gamma}$, o que indica uma probabilidade não desprezível de se obter vértices com graus muito maiores quando comparados ao restante da rede. (PASTOR-SATORRAS et al., 2015).

Para a construção de redes que sigam uma lei de potência, o trabalho de Barabasi e Albert (1999) desenvolve o conceito “anexação preferencial”. Baseado no trabalho de Price (1976) e a sua ideia de “vantagem acumulativa”, este modelo de crescimento provê a ideia de que vértices que possuem maior grau de entrada têm também maior probabilidade de receber uma nova conexão.

Observa-se que redes reais apresentam características na sua distribuição de grau semelhantes a uma lei de potência, em especial na “cauda” da distribuição.

2.1.4 Redes homogêneas e heterogêneas

Uma classificação importante para o contexto deste trabalho é sobre a visão de redes homogêneas ou heterogêneas. Esta classificação é feita basicamente em função da distribuição de grau $P(k)$ das redes. O estudo dos efeitos dinâmicos da difusão em redes complexas são seccionados entre estas duas categorizações, como pode ser visto no trabalho de Pastor-Satorras et al. (2015) para diferentes modelos, em especial para os epidemiológicos.

As redes são classificadas como homogêneas quando possuem uma distribuição de grau binomial ou de Poisson. Vê-se que a conectividade da rede está em torno do valor \bar{k} . Redes aleatórias e redes *Small World* atendem a esta característica específica.

As redes heterogêneas se assemelham a redes *Scale Free*, possuindo uma distribuição de grau $P(k)$ que seguem uma lei de potência. Podem ser observadas então regiões

com várias conexões e outras quase sem conexão.

2.1.5 Centralidade

Em redes complexas, a centralidade é uma característica topológica de grande importância. Este conceito procura quantificar a importância de um vértice da rede ou de suas ligações ante ao restante da rede. Existem diversas características que podem ser consideradas para a definição desta relevância e que são melhores aplicadas a diferentes sistemas complexos.

2.1.5.1 Grau

O grau é a medida mais intuitiva, caracterizando-se pela quantidade de conexões que um vértice possui. Seu escopo ganha especificidade para redes direcionadas, onde é medido o número de conexões de entrada e de saída dos vértices. Em síntese, na centralidade por grau, quanto maior for o número de conexões de um vértice, maior sua importância.

Em muitas redes reais, os vértices com alto grau são chamados de *hubs* e indicam uma tendência de anexação preferencial, como referenciado em [Barabasi \(2016\)](#).

Foi verificado que em diferentes estruturas de redes complexas, esta medida pode não conseguir caracterizar o vértice de maior relevância, o que levou pesquisadores à proposição de novas medidas.

2.1.5.2 Betweenness

Para a obtenção do *betweenness* mede-se o número de vezes que um determinado vértice está no trajeto mais curto entre dois outros vértices da rede. Quanto mais vezes este vértice compor diferentes trajetos, maior será sua importância na rede. Com isto, quanto maior for a necessidade deste vértice para a composição de menores caminhos, maior será sua centralidade.

Esta medida foi proposta no trabalho de [Freeman \(1978\)](#) e pode ser chamada como centralidade por intermediação.

2.1.5.3 Autovetor

A medida de centralidade de autovetor foi proposta em [Bonacich \(1987\)](#) e busca medir a importância de um vértice em função da importância de seus vizinhos. Espera-se que, mesmo que a centralidade de grau atribuída a um vértice tenha valor baixo, se os vizinhos forem importantes, fará com que esse vértice também seja importante. Matematicamente, a medida equivale ao autovetor associado ao maior autovalor da matriz de adjacência da rede.

2.1.5.4 Pagerank

O *Pagerank* foi proposto em [Page et al. \(1999\)](#) é utilizado para medir a importância de páginas na Internet considerando a qualidade e a quantidade dos links que direcionam a ela. Em síntese, a centralidade que se propaga aos vizinhos é proporcional ao número de arestas de saída. Vértices propagam apenas parte da centralidade adiante aos outros.

2.2 Epidemiologia Matemática

Durante mais de dois séculos, a modelagem matemática de epidemias foi alvo de estudos e pesquisas que envolviam diferentes áreas de conhecimento como matemática, biologia e ecologia. O trabalho de [Bernoulli \(1760\)](#), que estudou a varíola, é considerado o marco do estudo matemático de epidemias, embora apenas a partir da segunda metade do século XIX houvesse algum avanço no conhecimento médico sobre microrganismos e doenças infecciosas.

Anos de estudos e pesquisas tinham como objetivo um melhor entendimento sobre a incidência, propagação e persistência de epidemias ([ANDERSON; MAY, 1979](#)). A dedicação e interesse pela área permitiram a criação de um arcabouço conceitual capaz de auxiliar no entendimento e comportamento da disseminação de epidemias sob a perspectiva matemática. Livros clássicos como [Anderson e May \(1991\)](#), [Keeling e Rohani \(2008\)](#), [Brauer e Castillo-Chávez \(2012\)](#), [Bailey \(1975\)](#) e muitos outros apresentam e descrevem detalhadamente o cerne dessa área de estudo.

No contexto deste trabalho, que abrange conceitos das áreas computacional e de saúde, considera-se que epidemia é a alteração de características significativas de indivíduos de uma população através da interação entre estes e o meio. Esta mudança de característica é oriunda do contato com um patógeno ou com um indivíduo infectado, a considerar diferentes formas de transmissão dada a natureza da doença.

Alguns modelos matemáticos presumem que uma população pode ser dividida em compartimentos ou classes que classificam o indivíduo segundo seu estágio de doença ([ANDERSON; MAY, 1991](#)). Dentre algumas classes destacam-se:

Suscetível: indivíduo saudável que ainda não teve contato com o patógeno;

infectado: indivíduo que teve contato com o patógeno e que pode infectar outros indivíduos;

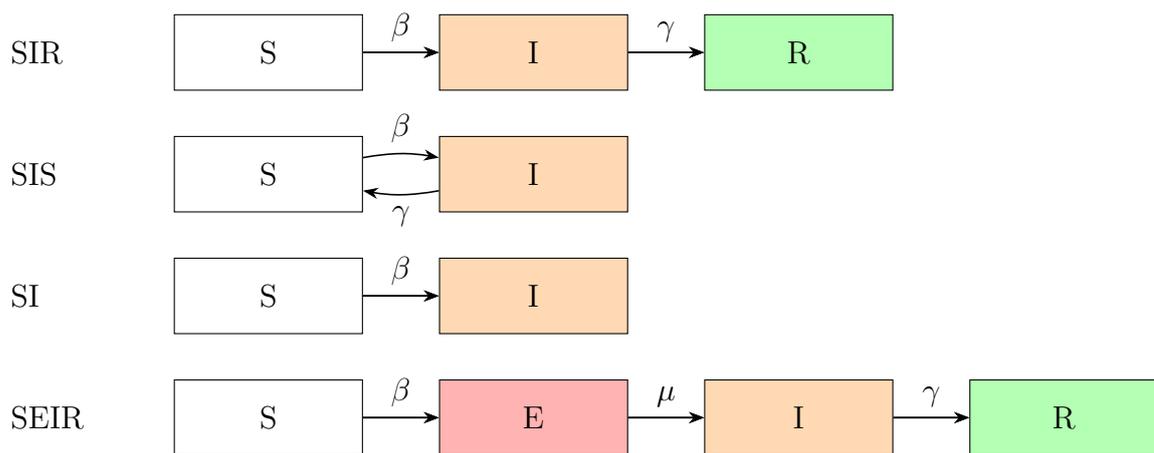
recuperado: antes infectado, indivíduo que se recuperou da doença;

expostos: indivíduo exposto a doença e que está em período de latência.

Esses compartimentos são representados a partir de sua primeira letra. Portanto, o grupos de suscetíveis, infectados, recuperados e expostos são denotados como S , I , R e E .

Os modelos buscam reproduzir a dinâmica de contágio de uma doença em uma população fixa e modelos mais simples ignoram aspectos demográficos, como migração e nascimentos (PASTOR-SATORRAS et al., 2015). A evolução do número de indivíduos em cada uma das classes é averiguada em função do tempo indicando a transição dos indivíduos entre os compartimentos. Parâmetros numéricos definem a “taxa de mudança” de indivíduos entre estados e normalmente são representados por letras gregas.

A nomenclatura desses modelos é dada em função da dinâmica de mudança da compartimentalização dos indivíduos. Pode-se citar o modelo SI (Suscetível-Infetado), SIS (Suscetível-Infetado-Suscetível), SIR (Suscetível-Infetado-Recuperado) e SEIR (Suscetível-Exposto-Infetado-Recuperado). A Figura 2 demonstra um diagrama desses principais modelos em decorrência de suas interações.



Fonte: Adaptado de (PASTOR-SATORRAS et al., 2015)

Figura 2 – Diagrama de representação de alguns modelos epidemiológicos

Modelos mais complexos e recentes podem requerer a definição de outros estados, como imune ou latente.

Esse trabalho dará enfoque ao modelo SIR por utilizá-lo como uma das principais ferramentas de estudo.

2.2.1 Modelo SIR determinístico

No modelo SIR, indivíduos transitam entre as classes suscetível, infectado e recuperado. A condição de imunidade é obtida logo que o indivíduo assume o estado de recuperado. Uma analogia válida é que esses são “removidos” da população, por ficarem incapazes de adquirir ou transmitir novamente a doença. Foi apresentado no trabalho de Kermack et al. (1927) e foi melhor detalhado em Dietz (1967).

Além da compartimentalização, o modelo assume a premissa de *mistura homogênea*, que espera que qualquer indivíduo tem a mesma chance de entrar em contato com qualquer outro indivíduo. Assim, o parâmetro β representa a probabilidade de infecção de um indivíduo e γ representa a probabilidade de recuperação. O modelo matemático determinístico pode ser descrito como na Equação 2.8.

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}\tag{2.8}$$

Uma informação importante, que decorre da Equação 2.8 e disposta em Keeling e Rohani (2008), é que se a fração inicial de indivíduos suscetíveis é menor que $\frac{\gamma}{\beta}$ tem-se que $\frac{dI}{dt} < 0$ e, com isso, a epidemia termina. No trabalho de Kermack et al. (1927) isso foi chamado de “*fenômeno limite*” em livre tradução. Por conseguinte, pode-se interpretar $\frac{\gamma}{\beta}$ é a taxa de remoção relativa. A inversão dessa razão é definida como *número básico de reprodução*, normalmente representado como R_0 . Na abordagem determinística, para que haja um surto epidêmico, basta que $R_0 > 1$.

Diz-se que R_0 representa um número médio de casos de infecção secundários frutos de um caso primário em uma população suscetível (KEELING; ROHANI, 2008).

2.2.2 O modelo SIR em redes complexas

A conexão entre os modelos epidemiológicos e redes complexas busca representar a natureza probabilística da disseminação de doenças e considera diversas características topológicas da rede. A intenção é representar melhor as interações aleatórias que acontecem em ambientes reais através de um modelo estocástico.

Em Barabasi (2016) descreve-se um modelo que considera um número típico de contatos $\langle k \rangle$ de um indivíduo onde a população tem N indivíduos. Esse valor $\langle k \rangle$ representa o grau médio da rede. A partir de um instante inicial $t = 0$ tem-se o número inicial de suscetíveis definidos por $S(0)$, de infectados $I(0)$ e de recuperados $R(0)$. Trabalha-se com frações da população, por isso pode-se representar cada um dos compartimentos em

razão do número de indivíduos da população como pode ser visto na Equação 2.9.

$$\begin{aligned} s(t) &= \frac{S(t)}{N} \\ i(t) &= \frac{I(t)}{N} \\ r(t) &= \frac{R(t)}{N} \end{aligned} \quad (2.9)$$

Ao assumir-se $S(0) = N$ e que existe a probabilidade $\frac{S(t)}{N}$ de um indivíduo infectado transmitir a infecção pode-se deduzir que o número de indivíduos suscetíveis atingíveis pode ser dado por $\langle k \rangle \frac{S(t)}{N}$.

Toda essa noção de proporcionalidade e estado inicial de uma doença, permite reescrever a Equação 2.8 com a substituição dos compartimentos em razão da população como pode ser visto na Equação 2.10. É possível notar que a função foi registrada resumidamente com a omissão da variável de tempo, ou seja, $s(t) = s$ por exemplo. Outro fator importante é que a equação passa a ser escrita em função do número de recuperados e infectados, pois assumiu-se que toda a população é suscetível no instante $t = 0$.

$$\begin{aligned} \frac{ds}{dt} &= -\beta \langle k \rangle i [1 - r - i] \\ \frac{di}{dt} &= -\gamma i + \beta \langle k \rangle i [1 - r - i] \\ \frac{dr}{dt} &= \gamma i \end{aligned} \quad (2.10)$$

O trabalho de [Pastor-Satorras et al. \(2015\)](#) descreve diferentes abordagens de modelos epidemiológicos sob a ótica de diversas áreas e devidamente aplicadas às redes complexas. É notória a forte correlação entre fenômenos físicos e o comportamento de epidemias.

O termo “*campo médio*”, em livre tradução, é extensamente utilizado e define a existência de independência entre os estados dos vértices e de seus vizinhos, assim como no modelo da Equação 2.10 que não considera a conectividade de cada indivíduo para a difusão.

Outra abordagem um pouco mais arrojada assume que existe em função do tempo as densidades parciais de infectados, recuperados e suscetíveis de nós de um mesmo grau k . Essas densidades parciais são denotadas como $\rho_k^I(t)$, $\rho_k^R(t)$ e $\rho_k^S(t)$ respectivamente. Assim a função destas variáveis no instante t pode ser descrita como nas Equações 2.11. Assume-se portanto que existe uma equivalência estatística entre os nós de grau k e suas

respectivas probabilidades.

$$\begin{aligned}\frac{\rho_k^I(t)}{dt} &= -\rho_k^I(t) + \lambda k \rho_k^S(t) \Gamma_k(t), \\ \frac{\rho_k^R(t)}{dt} &= \rho_k^I(t)\end{aligned}\tag{2.11}$$

onde

$$\begin{aligned}\rho_k^S(t) &= 1 - \rho_k^I(t) - \rho_k^R(t) \\ \Gamma_k(t) &= \sum_{k'} P(k'|k) \rho_{k'}^I(t).\end{aligned}\tag{2.12}$$

O termo $P(k'|k)$ representa a probabilidade de um vértice de grau k estar conectado a outro de grau k' . Essa abordagem é chamada, em livre tradução, de SIR de campo médio baseado em grau e é detalhada na importante revisão de [Pastor-Satorras et al. \(2015\)](#).

O SIR aplicado a este trabalho considera cada um dos vértices e suas condições, sem generalização da probabilidade por grau por exemplo, como no modelo destacado acima. O método individual e baseado em pares, discutido em [Peixoto \(2020\)](#), considera os vértices e suas conexões para determinar a mudança de estado indivíduo a indivíduo. Assim, nós suscetíveis podem ser contaminados por vizinhos infectados. Todo indivíduo infectado pode ainda se recuperar eventualmente.

Cada vértice i possui um estado s_i em um instante t , onde sua classe é definida conforme a Equação 2.13

$$\begin{cases} s_i(t) = 0 & \text{se } i \text{ está suscetível} \\ s_i(t) = 1 & \text{se } i \text{ está infectado} \\ s_i(t) = 2 & \text{se } i \text{ está recuperado} \end{cases}\tag{2.13}$$

Um vértice i tem seu estado alterado de $s_i(t) = 0$ para o estado $s_i(t+1) = 1$ com probabilidade vista na Equação 2.14. Já a mudança de $s_i(t) = 1$ para $s_i(t+1) = 2$ ocorre com a probabilidade γ_i .

$$(1 - h_i) \left[1 - \prod_j (1 - \beta_{ij})^{A_{ij} \delta_{s_j(t), 1}} \right] + h_i\tag{2.14}$$

onde:

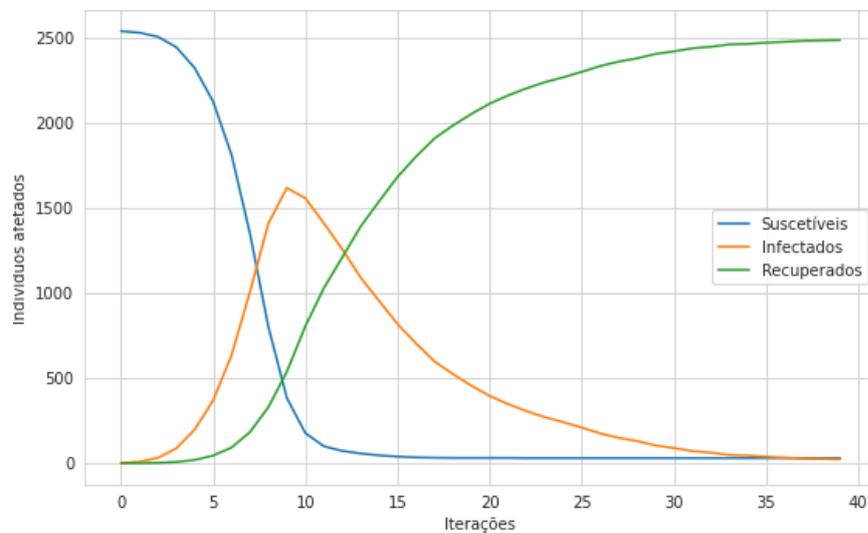
β_{ij} = a probabilidade de transmissão para cada aresta da rede

A_{ij} = posição na matriz de adjacência A que indica se há conexão entre os vértices i e j

h_i = representa a taxa de infecção espontânea do vértice i

A discussão do termo $\delta_{s_j(t),1}$ foi intencionalmente omitida para uma explicação mais detalhada. Ele representa o delta de Kronecker, uma notação matemática que retornará o valor 1 se os dois índices forem iguais e o valor 0 caso contrário. Na Equação 2.14 o termo busca indicar se o estado do nó j é 1, ou seja, infectado.

É possível verificar que a fração de indivíduos infectados cresce exponencialmente e que, após atingir o pico, $i(t \rightarrow \infty) = 0$. O número de suscetíveis possui o mesmo comportamento desde o início da série histórica. A Figura 3 ilustra o comportamento de uma epidemia aplicada a uma rede complexa com 2530 vértices onde se infectou o nó de maior grau.



Fonte: O autor

Figura 3 – Comportamento epidêmico do modelo SIR aplicado em uma rede complexa. Utilizou-se $\beta = 0.2857$ e $\gamma = 0.1428$

No contexto deste trabalho, características como sentido das arestas e peso serão generalizados, pois o foco é apenas considerar o contato entre os indivíduos.

2.3 Algoritmos genéticos

A busca por soluções de problemas NP-difíceis¹ ou problemas de classe P que não admitem um tempo flexível $\mathcal{O}(n^x)$ é um dos principais desafios da computação. Busca-se portanto algoritmos não-determinísticos capazes de encontrar uma solução cada vez melhor a partir de um espaço de soluções possíveis. A busca por soluções satisfatórias para esses problemas desencadeou no estudo de heurísticas e metaheurísticas. A ideia é prover soluções sem garantia de otimalidade.

Segundo Blum e Roli (2003), algoritmos metaheurísticos e heurísticos são técnicas que abordam desde procedimentos de busca local simples até processos de aprendizagem

¹ Problemas que não possuem nenhum algoritmo de ordem polinomial capaz de prover uma solução determinística

complexos. É importante que haja um bom equilíbrio entre duas características: diversificação e intensificação. A diversificação remete à exploração de outras regiões do espaço de busca enquanto a intensificação busca a exploração da região utilizando estruturas experimentadas e que levam a soluções de boa qualidade.

Esses algoritmos são classificados em duas classes diferentes. A primeira considera apenas algoritmos que trabalham com apenas uma solução por vez. Outra considera metaheurísticas que trabalham com múltiplas soluções ou uma população de soluções. Essas classes compartilham um bom número de conceitos e possuem aplicações muitas vezes distintas. Existem ainda implementações híbridas que utilizam das duas classes para a obtenção de melhores resultados.

No contexto deste trabalho discute-se a aplicação de algoritmos genéticos, que pertence a classe de metaheurísticas populacionais, ou seja, que trabalha com múltiplas soluções.

Segundo [Goldberg et al. \(1981\)](#), os algoritmos genéticos foram propostos por [Holland](#) em 1975 e buscou abstrair o comportamentos observados em sistemas naturais criando algoritmos capazes de simular estes comportamentos artificialmente.

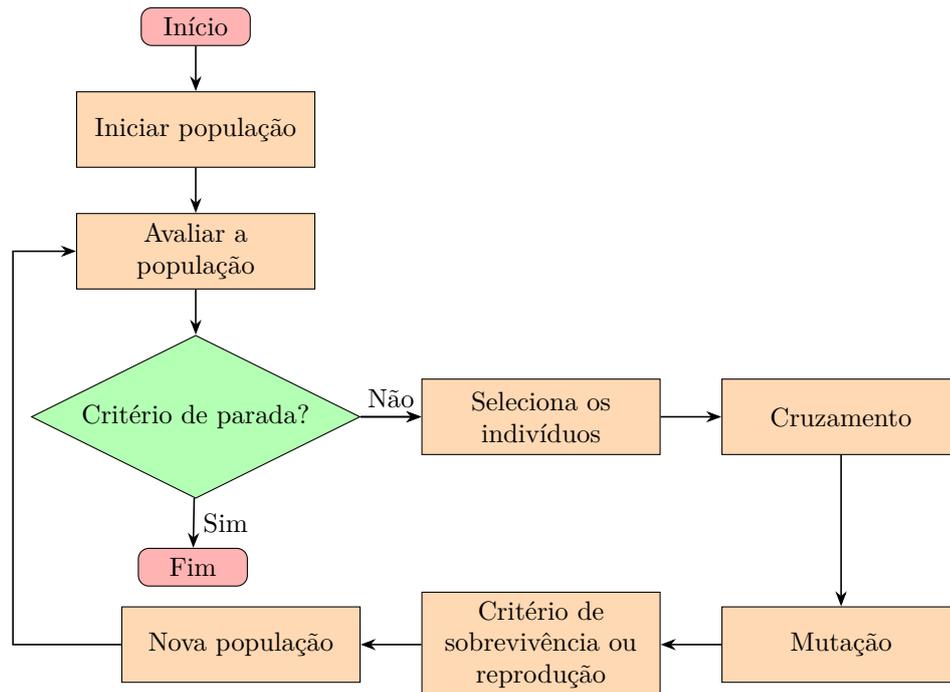
Algoritmo genético, ou AG, faz parte da família de algoritmos evolutivos aplicados a problemas de busca e otimização. Este conjunto busca reproduzir diferentes mecanismos biológicos inspirados em fenômenos que ocorrem na natureza ([BACK, 1996](#)). Inspiram-se em técnicas de biologia evolutiva e buscam reproduzir o comportamento da seleção natural e da genética. Se utilizam destas técnicas para realizar uma busca paralela e estruturada. Foca no reforço de buscas em pontos de melhor aptidão através da característica da seleção natural, onde indivíduos mais adaptados ao meio tendem a perpetuar suas características, o que minimiza o fator aleatoriedade. O método usualmente utilizado foi proposto por [Holland \(1975\)](#).

Um dos principais diferenciais dessa estratégia é a de se trabalhar com uma população de soluções candidatas. Outra característica importante é que a maioria das regras de transições dentre os seus diversos passos são realizados probabilisticamente. A [Figura 4](#) ilustra a dinâmica de etapas que um AG deve realizar.

2.3.1 Conceitos gerais

Problemas de otimização são definidos a partir de:

- variáveis $X = \{x_1, x_2, x_3, \dots, x_n\}$;
- o domínio de cada variável;
- restrições entre variáveis;



Fonte: O autor

Figura 4 – Diagrama de representação de um algoritmo genético

- uma função objetivo que será otimizada;

O conjunto de todas as soluções W determina o espaço de busca da solução.

Uma solução $w^* \in W$ é ótima se e somente se para toda solução $w \in W$ se verifica $f(w^*) \geq f(w)$ em problemas de maximização ou se verifica $f(w^*) \leq f(w)$ em problemas de minimização.

2.3.2 Representação de soluções

A representação das soluções é uma importante escolha a se tomar na modelagem de qualquer heurística. Esta escolha impacta diretamente na dinâmica de formação e avaliação das soluções. Uma má escolha do método de representação pode adicionar custos computacionais à modelagem proposta e por consequência soluções distantes dos valores ótimos. Apresenta-se duas formas de representação:

Linear strings de símbolos de um alfabeto conhecido. Variáveis de decisão denotam a presença ou ausência de um elemento (codificação binária)

Não linear geralmente baseada em estrutura de grafos

É necessário que a representação usada consiga representar devidamente todas as soluções possíveis do problema (completude); deve sempre existir um caminho de busca entre quaisquer duas soluções (conexidade) e ser de fácil manipulação pelos operadores de busca (eficiência).

2.3.3 Função objetivo

A função objetivo ou de aptidão associa uma solução do espaço de busca a um valor determinístico que descreve a aptidão ou qualidade da solução. Essa característica permite a ordenação ou classificação de soluções.

Em Talbi (2009) diz-se que a função objetivo é direta quando retorna um valor quantitativo que basta para sua avaliação. Pode ser ainda classificada como uma função guia ao avaliar qualitativamente a efetividade da função, direcionando para uma melhor solução.

2.3.4 Tratamento de restrições

Alguns problemas podem apresentar restrições que inviabilizam ou dificultam a busca por soluções. Com isso, pode-se elaborar estratégias que utilizem da representação e da função objetivo como mecanismo de ajuste.

A rejeição é uma estratégia utilizada que trabalha com o descarte de soluções inviáveis. Implementações mais arrojadas consideram um determinado nível de inviabilidade, para que se evite qualquer convergência prematura.

A penalização não realiza qualquer descarte de soluções, mas busca penalizar aquelas que violam alguma restrição. A penalização é normalmente linear, algo como $f'(w) = f(w) + \alpha * c(w)$ onde $c(w)$ é o custo da violação da restrição e α é um peso agregado.

A estratégia de reparação aplica um algoritmo corretivo para que se corrija soluções inviáveis. É desejável alguma eficiência computacional deste algoritmo para que se mantenha a viabilidade do método.

Outra estratégia importante e óbvia é a da preservação, que trabalha apenas no espaço de busca de soluções viáveis e que atendam todas as restrições do problema proposto.

Em Talbi (2009) vê-se detalhadamente o uso e outras considerações sobre técnicas de tratamento de restrições.

2.3.5 Sujeitos no AG

A modelagem define os seguintes sujeitos importantes para o entendimento da dinâmica:

Genes são as características que formam o indivíduo.

Cromossomo ou indivíduo conjunto de genes ou características que formam uma possível solução do problema.

População conjunto de diversos indivíduos ou cromossomos, que são candidatos a solução do problema.

Geração cada etapa ou avanço de passo do AG que produzem uma nova população ou uma alteração expressiva desta.

2.3.6 População inicial

A população inicial de um AG é parte essencial do trabalho e forma um conjunto de indivíduos candidatos à solução do problema para o início do trabalho de otimização. É desejável que seja diversificada o bastante para evitar qualquer convergência prematura. Sua criação pode ser realizada de forma aleatória ou partir de alguma heurística construtiva. Anseia-se que a solução do AG não seja determinada exclusivamente pela formação da população inicial.

2.3.7 A seleção

A seleção é uma das fases utilizadas por AG's que buscam selecionar quais indivíduos da população serão escolhidos para a aplicação dos operadores genéticos e que formarão uma nova população.

A seleção proporcional, ou por roleta, foi proposta por [Holland \(1975\)](#) e usa o resultado da função objetivo dos indivíduos normalizada para criar uma estrutura fatiada. O sorteio de um valor entre zero (0) e um (1) define a escolha da fração que é contida por um único indivíduo da população.

A seleção por torneio foi proposta em [Goldberg et al. \(1981\)](#) e é feita a partir da escolha aleatória de x indivíduos, selecionando-se o melhor deles a partir de seu valor da função objetivo.

Já a seleção por truncamento define um limiar dos melhores indivíduos candidatos a serem selecionados. Deste grupo, a seleção do indivíduo é feita aleatoriamente. Esta estratégia foi apresentada no trabalho de [Mühlenbein e Schlierkamp-Voosen \(1993\)](#).

O trabalho de [Blickle e Thiele \(1995\)](#) realizou um extenso comparativo destas e outras estratégias e suas aplicações.

2.3.8 Cruzamento

O cruzamento é a fase posterior à seleção e que produz descendentes a partir dos indivíduos selecionados. Seu principal objetivo é intensificar a busca por soluções através da utilização de partes dos indivíduos já avaliados. Esse processo se repete até que se tenha o número de indivíduos necessários para o avanço de uma geração.

A estratégia mais utilizada e a considerada neste trabalho é a de n-pontos, que reparte os cromossomos em grupos de genes que são mesclados. Espera-se que os novos indivíduos produzidos possam ser melhor avaliados por conterem a “carga genética” dos pais selecionados. O trabalho de [Lim et al. \(2017\)](#) descreve esta e inúmeras outras estratégias de cruzamento.

A Figura 5 representa o cruzamento de dois pontos entre dois indivíduos e que forma outros dois.

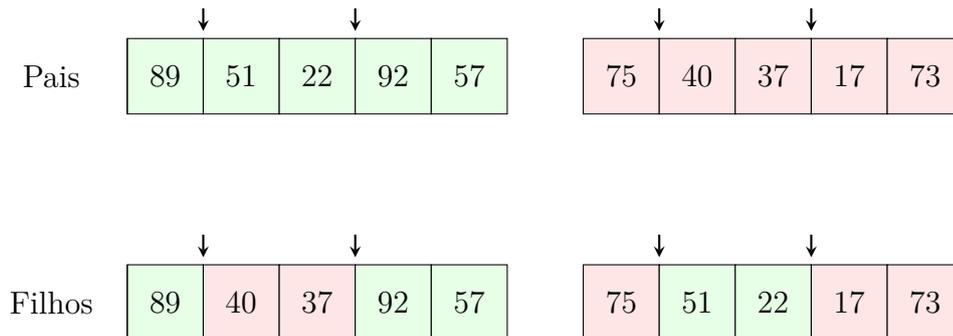


Figura 5 – Cruzamento de dois pontos em cromossomos com 5 genes.

2.3.9 Mutação

A fase de mutação busca gerar diversidade de soluções e garante a possibilidade de melhor varredura do espaço de busca. Permite a mudança normalmente aleatória de alguns genes com uma probabilidade baixa de ocorrência.

Neste trabalho o enfoque será no método aleatório que testa para cada gene de cada indivíduo alguma probabilidade de mutação. O gene selecionado possui o seu valor alterado randomicamente, como na Figura 6. O trabalho de [Lim et al. \(2017\)](#) também discute diversas estratégias de mutação.

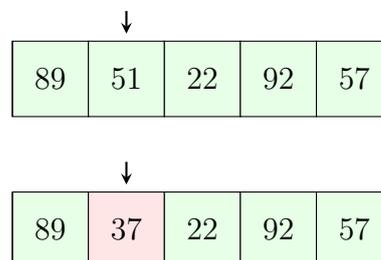


Figura 6 – Mutação de um gene definido pela probabilidade de mutação

2.3.10 Critérios de sobrevivência ou reprodução

Até este passo, o AG criou uma nova população de indivíduos com um tamanho específico e que passou por uma série de fases que geraram diversidade e alguma intensificação da busca.

O AG dito geracional substitui toda a população a cada geração. Pode ser aplicado um mecanismo elitista, que busca manter o melhor indivíduo da geração anterior para que se garanta a permanência de carga genética de qualidade para a próxima geração. O elitismo garante a tendência monotônica de busca de uma solução ótima; crescente em casos de maximização e decrescente em casos de minimização.

A versão “steady-state” do AG, ou SSAG, não utiliza o conceito de gerações por realizar a competição entre pais e filhos pela sobrevivência na população. Com isso apenas uma fração da população é gerada por vez e deve substituir outros da população anterior segundo algum critério. O mais comum é a substituição dos piores indivíduos da população pelos novos indivíduos gerados desde que estes possuam maior aptidão. Nos trabalhos de [Dréo et al. \(2006\)](#), [Syswerda \(1991\)](#) e [Loureiro \(2016\)](#) diz-se que esse número de indivíduos gerados costumeiramente é de um ou dois indivíduos. Outras abordagens assumem a substituição de um número maior de indivíduos, como em [Babu \(2013\)](#).

O termo “*gap*” foi apresentado no trabalho de [Jong \(1975\)](#) e tem o objetivo de controlar a fração da população que será substituída a cada geração/iteração. O trabalho de [Jong e Sarma \(1993\)](#) realizou uma extensa análise dos efeitos da variação deste “*gap*” e seus efeitos.

3 A metodologia GASIR

A metodologia experimental proposta neste Capítulo baseia-se no uso de algoritmos genéticos para a escolha de uma combinação de indivíduos para imunização em uma população representada através de uma rede complexa. Espera-se que o conjunto de vacinados sugeridos pela metodologia seja capaz de minimizar os impactos da disseminação de uma doença, em cenários cujo os primeiros indivíduos infectados são desconhecidos. Os conjuntos de indivíduos candidatos a solução são avaliados “in silico” utilizando-se de simulações epidêmicas com o modelo epidemiológico SIR em busca de se minimizar o número de indivíduos da rede afetados por uma doença considerada.

O AG propõe como solução um conjunto de vértices ideal e que considera os indivíduos infectados e os parâmetros da doença analisada. Para que se cumpra o objetivo da metodologia de destacar os melhores indivíduos a se imunizar independentemente da origem da epidemia, são realizadas múltiplas execuções do AG. Este passo adiciona diferentes perspectivas da disseminação da doença na rede e que, sumarizadas através de critérios que serão discutidos, destacam os vértices mais importantes para imunização na rede.

A metodologia GASIR, acrônimo de *Genetic Algorithm for Susceptible, Infected and Recovered*, recebeu esse nome por ser baseada na estrutura proposta no trabalho de [Rodrigues \(2018\)](#).

Nas seções posteriores será discutida as inúmeras características da metodologia, que utiliza uma função estocástica para avaliação de soluções de um algoritmo de otimização.

Um destaque importante no contexto deste capítulo é que o termo indivíduo remeterá sempre a um vértice da rede. Qualquer referência à palavra indivíduo como sujeito de um AG será realizada pelo termo cromossomo.

3.1 Definições, ferramental e o cenário experimental

A metodologia foi implementada com a utilização da linguagem de programação *Python*. Para realização das simulações epidêmicas e para o trabalho com redes complexas, decidiu-se por utilizar o *framework graph-tool*¹, que tem como uma de suas premissas o compromisso com a performance computacional, seria a melhor opção. O *framework* provê um módulo para a linguagem *Python*, mas que implementa a maioria dos algoritmos em

¹ <https://graph-tool.skewed.de/>

C/C++ utilizando-se da biblioteca *Boost.Python*². Segundo (PEIXOTO, 2014) a utilização desse ferramental permite que esses algoritmos possuam desempenho computacional comparável àqueles desenvolvidos puramente em *C/C++*. Outro fator que corrobora com o acerto dessa escolha é que o *framework* implementa diversos modelos matemáticos de processos de difusão em redes complexas nativamente, dentre eles o SIR.

Para justificar e ilustrar diversas tomadas de decisão adotadas para a proposição da metodologia, faz-se necessário a apresentação do ambiente de experimentos utilizados e outras definições que facilitarão a discussão.

Foram escolhidas duas redes complexas pequenas, de características topológicas diferentes para a realização de testes e para a proposição da metodologia. Como o problema proposto considera apenas o contato entre indivíduos como essencial, características como direção e peso das arestas existentes na organização das redes foram desconsideradas. Outra correção aplicada é que arestas paralelas oriundas de relacionamentos recíprocos foram removidas.

A Tabela 2 descreve as características das redes após os ajustes relatados.

Tabela 2 – Dados gerais das redes utilizadas para simulações. Considere que as redes possuem $|V|$ vértices, $|E|$ arestas, \bar{k} grau médio, k_{max} de grau máximo e k_{min} de grau mínimo. As colunas c e r representam o coeficiente de *clustering* e assortatividade respectivamente.

Rede	$ V $	$ E $	\bar{k}	k_{max}	k_{min}	c	r
moreno highschool	70	274	7.83	19	2	0.40	0.01
moreno innovation	241	929	7.71	28	1	0.25	-0.01

Definiu-se um *experimento base* que representa um cenário ótimo de imunização para auxiliar no ajuste de parâmetros da metodologia. Neste experimento, considera-se que após a infecção do vértice 21 da rede *moreno highschool* o resultado ótimo de vacinação é determinado pela escolha dos seus 8 vizinhos diretos.

Para a facilitar a retomada da discussão, alguns termos foram definidos conforme a seguir:

Contenção direta: representa a escolha de todos os vizinhos diretos de um conjunto de vértices infectados.

Estabilidade vertical: averigua a capacidade do modelo de otimização em retornar como solução conjuntos de vértices semelhantes para vacinação após sucessivas execuções da metodologia com os mesmo parâmetros.

Estabilidade horizontal: averigua a capacidade do modelo de otimização em manter alguns vértices como solução dado o aumento do número de vacinas disponíveis.

² <https://www.boost.org>

A terminologia de “estabilidade” definida é utilizada na discussão deste trabalho para se referenciar a capacidade da metodologia de prover soluções semelhantes em diferentes execuções. O uso de um método de otimização junto a uma função estocástica pode gerar resultados bastante aleatórios, característica indesejável para a solução do problema proposto.

3.2 Modelagem da metodologia

Para a utilização do AG como recurso de otimização modelou-se que cada gene seria representado pelo índice único que registra um vértice candidato a vacinação. O tamanho desse conjunto de genes ou cromossomo define portanto o número de vacinas disponíveis.

A Figura 7 ilustra a estrutura utilizada. Dada uma rede complexa $G = (V, E)$ tal que $V = \{0, 1, 2, \dots, 100\}$ vê-se a representação de um AG com três(3) gerações, população de quatro (4) cromossomos com quatro (4) genes cada.

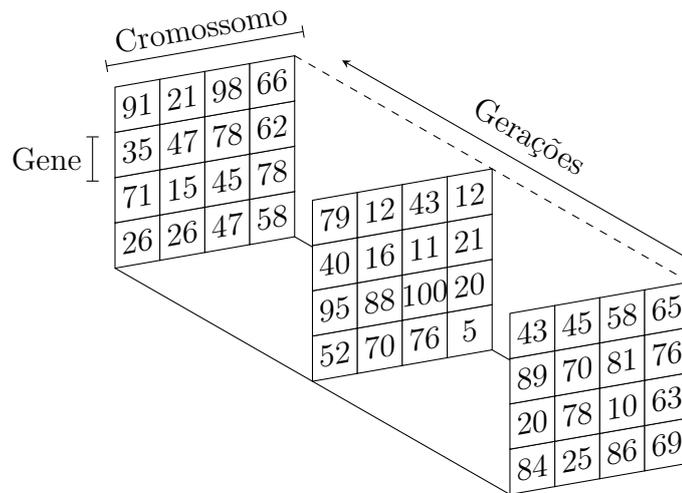


Figura 7 – Representação gráfica do AG.

A avaliação dos cromossomos é realizada por um modelo SIR estocástico, baseado em indivíduos, que considera as interações entre os vértices da rede complexa para simular difusão da epidemia. Um vértice infectado tem a probabilidade β de transmitir a doença a um vizinho suscetível. Todos os infectados têm a probabilidade γ de se recuperarem. Ressalta-se que um vértice suscetível tem probabilidade de infecção diretamente proporcional ao número de vizinhos também infectados, que é dada pela Equação 2.14.

O modelo SIR utilizado neste trabalho para as simulações em redes complexas provê parâmetros para que se defina uma probabilidade de infecção distinta para cada aresta entre dois vértices quaisquer i e j da rede, dada pelo parâmetro β_{ij} ; a probabilidade de recuperação distinta para cada vértice i , dada por γ_i ; e a probabilidade de infecção

espontânea de cada vértice, dada por h_i . Todos os experimentos aplicados basearam-se em parâmetros epidêmicos estimados para uma epidemia de *Influenza* e obtidos através de informações dos trabalhos de Bresee et al. (2011) e Coburn et al. (2009). Assim, a taxa de infecção β assumida foi de 0.2857 em todas as arestas das redes, a taxa de recuperação γ de 0.1428 para todos os vértices e a atribuição da taxa de infecção espontânea $h_i = 0$.

Define-se um grupo de vértices que permanece infectado durante todas as gerações do AG e que não varia no decorrer das gerações. Para a avaliação de cada cromossomo, move-se todos os indivíduos que a ele pertencem para o compartimento de recuperados. Todos os membros desse compartimento são isentos da possibilidade de infectar quaisquer outros vértices, característica que representa a imunização provida por uma vacina. O restante dos indivíduos fica no compartimento de suscetíveis, em condição de vulnerabilidade.

Finalizada a alocação dos indivíduos, executa-se a simulação da epidemia. A avaliação do cromossomo é definida pela contagem de indivíduos cujo estado é infectado ou recuperado após o fim do processo epidêmico. A meta do algoritmo é minimizar o valor dessa função objetivo. Em um seção própria será discutida todo o mecanismo considerado para a avaliação dos cromossomos do AG.

Após a fase de avaliação, são realizadas as operações de seleção, cruzamento e mutação dos cromossomos para a formação de uma população que avançará para a próxima geração. O critério de parada do AG é definido por um número fixo de gerações.

A seleção dos indivíduos a serem vacinados segue algumas premissas. Para epidemias cuja origem da infecção é conhecida considere as diretrizes abaixo.

- A metodologia ajusta a alocação das vacinas segundo os parâmetros da doença avaliada e o número de doses disponíveis. Assim, a mudança de quaisquer desses parâmetros determina a necessidade de uma nova execução da metodologia.
- A solução dada por uma execução do AG determina um grupo de indivíduos a serem imunizados para o conjunto de infectados utilizados no AG.
- Ao final de uma execução do AG, retira-se os 10% melhores cromossomos avaliados na última geração do AG.
- Os vértices mais frequentes nesses cromossomos, selecionados até o limite de doses disponíveis, determinam o grupo de vacinados ideal.

A escolha dos vértices para uma epidemia de origem desconhecida difere da anterior no número de execuções do AG. Múltiplas execuções adicionam diferentes perspectivas de espalhamento da doença na rede. Desta forma definiu-se as seguintes premissas: ’

- São necessárias 100 execuções do AG, que se diferenciam pelo grupo de infectados inicial.

- Os grupos de infectados tem tamanho igual a 5% da população.
- De cada execução do AG são retirados também 10% dos melhores cromossomos.
- A solução final é dada pela frequência acumulada dos vértices contidos nos cromossomos extraídos de todas as execuções do AG, selecionados até o limite de doses disponíveis.

Os experimentos assumem o conceito de vacinação passiva, ou seja, os indivíduos são vacinados em momento anterior ao início da epidemia.

Em seu trabalho, [Rodrigues \(2018\)](#) considera que a solução para qualquer epidemia, mesmo com a origem da infecção determinada, depende de múltiplas execuções do AG e que a partir do histograma formado pelos vértices que pertencem aos 10% melhores cromossomos da última geração de todas as execuções do AG definiria o conjunto de vacinados ideal.

A presente proposta parte da hipótese de que uma única execução do AG basta para a proposição do conjunto de vacinados ideal para um grupo de infectados qualquer e que as sucessivas execuções do AG são necessárias para a adição de novas perspectivas de infecção para a análise dos vértices mais importantes.

Outras operações genéticas e ajustes da função objetivo são também consideradas na metodologia atual, que avalia ainda o nível de aleatoriedade que as soluções propostas apresentam.

Os valores de parâmetros do AG propostos por [Rodrigues \(2018\)](#) são utilizados como marco inicial para a discussão desta metodologia.

3.3 A função objetivo

A função objetivo planejada para aplicação no GASIR é dada pela aplicação do modelo SIR estocástico e que mensura o número de indivíduos infectados ou recuperados após a simulação da epidemia. Esse passo representa grande parte do custo computacional da metodologia, visto o extenso número de avaliações necessárias para que se produza alguma solução.

O comportamento probabilístico do modelo epidemiológico gera um efeito dicotômico e indesejável a algoritmos de otimização. Inúmeras simulações poderiam tanto terminar precocemente quanto tender à infecção de todos os vértices da rede. Para ilustrar esse comportamento as Figuras 8 e 9 representam visualmente duas simulações onde um mesmo vértice infectado foi capaz tanto de disseminar a doença para toda a rede quanto de sequer realizar algum contágio.

Os parâmetros do SIR utilizados nos experimentos indicam um surto epidêmico, pois $\beta > \gamma$, ou seja, a probabilidade de infecção é superior a probabilidade de recuperação. Com isso inúmeros conjuntos de vacinados serão mal avaliados graças à tendência de se infectar todos os vértices da rede. Exceções a esse comportamento ocorrerão esporadicamente, especialmente se forem vacinados todos os indivíduos vizinhos aos infectados ou se a vacinação definir uma falha topológica, através da quebra da componente gigante da rede complexa. Esse comportamento é potencialmente problemático por dificultar a captura de uma melhora gradativa das soluções propostas, recurso essencial para algoritmos de otimização.

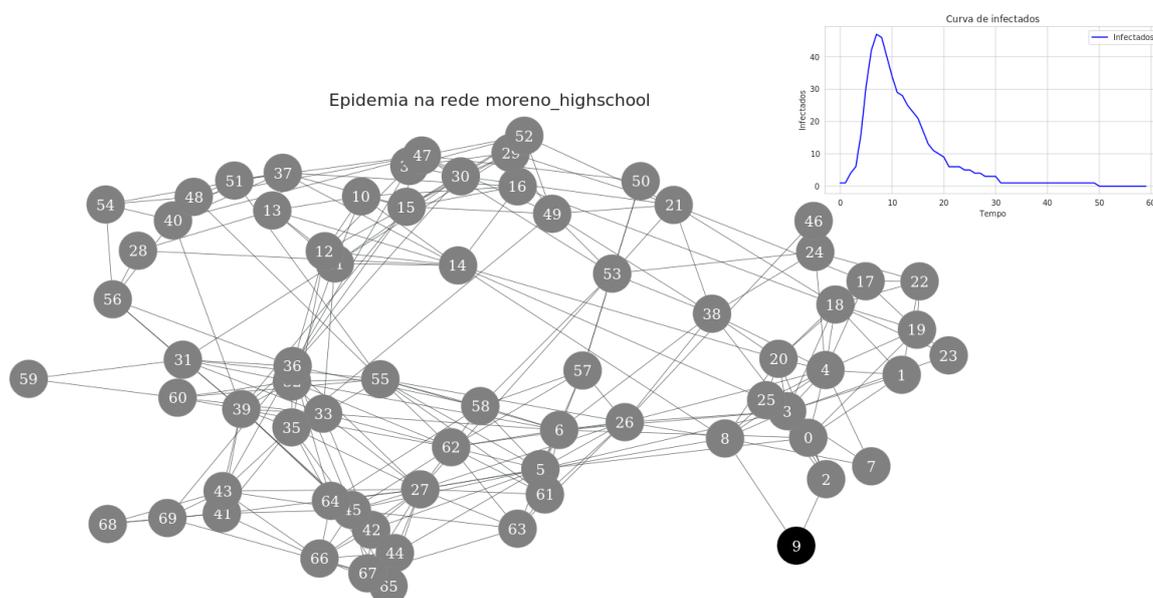


Figura 8 – Epidemia completa. Vértices pretos foram infectados; cinzas recuperados.

Averiguou-se a possibilidade de se considerar outras medidas da curva epidêmica **de infectados** como critério de avaliação dos cromossomos. Cada uma das alternativas foi verificada através de experimentos isolados que tinham como linha de base as diretrizes do *experimento base* citado na Seção 3.1.

A maioria das medidas consideradas levam em consideração a série histórica de infectados da epidemia simulada. Tornou-se necessário que, a cada instante/iteração do modelo SIR fosse computado o compartimento de cada um dos vértices da rede para que se realizasse a soma do número de infectados no instante t avaliado. Essa estratégia adicionou um custo computacional alto para as inúmeras simulações do SIR e não geraram qualquer indício de estabilidade do método. Todas as medidas avaliadas são listadas a seguir:

- Integral do polinômio interpolador (regra dos trapézios) aplicada a curva de infectados.
- Pico de infectados da epidemia simulada.

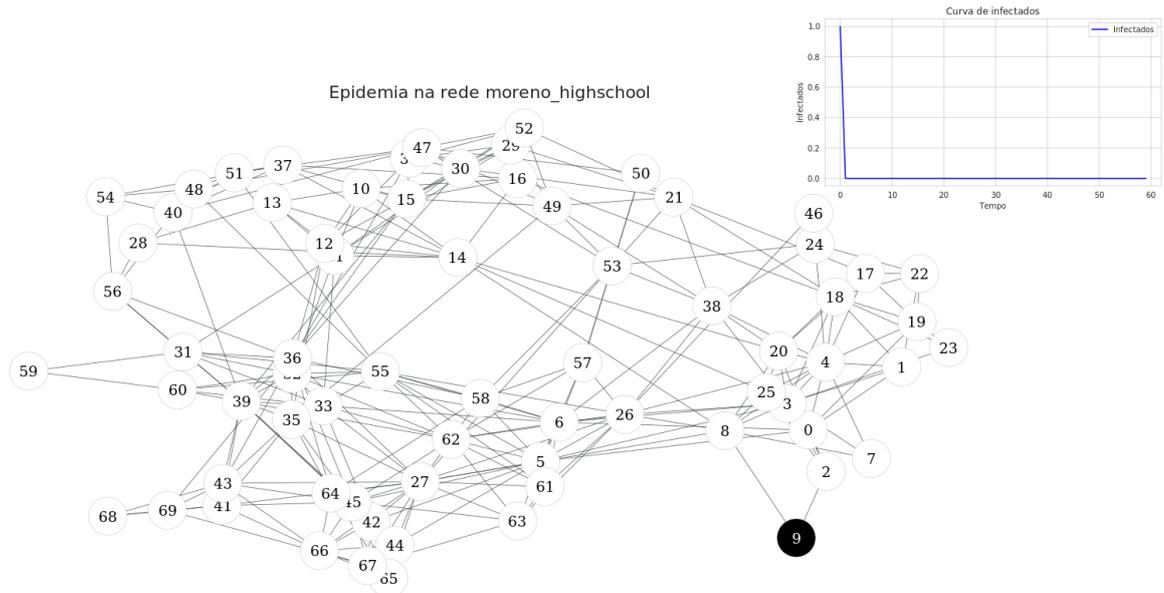


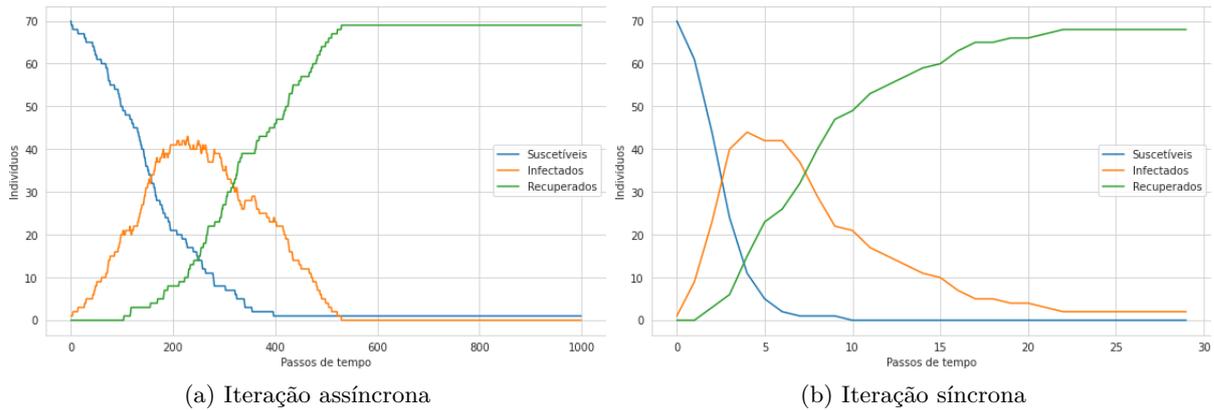
Figura 9 – Epidemia com fim precoce. Vértices pretos foram infectados; brancos suscetíveis.

- Coeficiente de inclinação da reta que parte do início da série histórica de infectados até o par ordenado que representa o instante/pico de infectados da série analisada.
- Análise comparativa de medidas de estatística descritiva para a curva de infectados. Considerou-se a curtose e a assimetria.

Partiu-se para investigação da dinâmica do modelo SIR. Observou-se dois mecanismos diferentes de iterações providos pelo *graph-tool* para a verificação da probabilidade de mudança de estado dos indivíduos da rede nas simulações e que possuíam objetivos parecidos. São permitidas iterações síncronas ou assíncronas no modelo. Esta averíua a probabilidade de recuperação e transmissão utilizando-se apenas de um vértice infectado da rede por iteração; aquela realiza atividade semelhante, porém verifica todos os vértices no instante. A Figura 10, que simula na rede *moreno_highschool* o modelo SIR infectando-se inicialmente o vértice de maior grau, representa as duas diferentes iterações que, embora estocásticas, possuem certa semelhança comportamental na curva de infectados. A principal diferença é caracterizada pelo número de iterações utilizadas, conforme registro no eixo das abscissas.

Viu-se potenciais problemas na utilização de iterações síncronas por essa atingir numericamente o pico da epidemia com poucas iterações. A expectativa é que o maior número de iterações utilizadas no mecanismo assíncrono facilite a realização de paralelos entre diferentes conjuntos de vacinados.

Assim, a coleta do número de afetados em instante/iteração anterior ao pico na curva de infectados captura o crescimento do número de casos, o que inibe a necessidade de se averiguar a alocação dos indivíduos a cada instante/iteração do SIR.



Fonte: O autor

Figura 10 – Simulações do modelo SIR com iteração assíncrona e síncrona

Como forma de se mitigar a probabilidade de fim precoce da disseminação da doença, o caminho mais óbvio é realizar múltiplas simulações da epidemia para o mesmo conjunto de vacinados, desconsiderando-se aquelas que finalizassem abruptamente. É sabido que tal estratégia adiciona um custo computacional, mas que pode garantir o melhor funcionamento do modelo.

Resumidamente, o processo de avaliação dos cromossomos do AG foi definido conforme as diretivas a seguir. Esta dinâmica procura minimizar as limitações observadas na utilização de uma função objetivo baseada em um modelo SIR estocástico.

1. Utilização de iterações assíncronas no SIR, em número de vezes igual ao número de vértices da rede. Verificou-se que este número de iterações é normalmente anterior ao pico de infectados da epidemia.
2. A realização de múltiplas execuções do SIR a cada avaliação de um cromossomo do AG, como método de reparação aplicado a função objetivo para que se evite o fim precoce do espalhamento da doença. Optou-se por dez sucessivas execuções da avaliação.
3. O valor de avaliação de cada cromossomo é dado pelo resultado que registra o maior número de infectados e recuperados (somados) dentre as dez sucessivas execuções e que registra a pior das simulações.

Todas diretivas acima dispostas foram definidas através de análises “*in silico*” de execuções amostrais da metodologia aplicadas no *experimento base*.

Matematicamente, considere uma rede complexa dada por $G(V, E)$, onde V representa o conjunto de vértices da rede e E o conjunto de arestas. Para a avaliação de um cromossomo do AG, simula-se a epidemia com o número de iterações igual n ou $|V|$. Tem-se portanto um conjunto-solução $S = \{s_0, s_1, s_2, \dots, s_{n-1}\}$ onde cada valor do conjunto, na forma s_i , representa o estado do vértice i com os possíveis valores registrados

na Equação 2.13. É possível afirmar portanto que a cardinalidade do conjunto S é igual o número de vértices da rede.

Considere ainda que w_j representa a j -ésima simulação de epidemia para o mesmo grupo de infectados e vacinados e que registra a soma de indivíduos infectados e recuperados ao fim da simulação. Tem-se portanto que o resultado de uma única avaliação de um cromossomo é definido pela Equação 3.1.

$$w_j = \sum_{i=0}^{n-1} [s_i = 1] + \sum_{i=0}^{n-1} [s_i = 2] \quad (3.1)$$

onde:

- n = número de vértices da rede estudada ou $|V|$;
- i = índice único que representa um vértice na rede;
- s_i = estado final do vértice após a simulação da epidemia.

Com isso, a avaliação final do cromossomo poderá ser representada conforme a Equação 3.2.

$$fit = \max\{w_j \mid j = 1, 2, \dots, 10\} \quad (3.2)$$

3.4 O algoritmo genético

No algoritmo de otimização da metodologia GASIR assumiu-se diversas características de um AG clássico. Dentre as premissas e operações adotadas, destaca-se:

- O critério de parada do AG é definido por um número fixo de gerações;
- a seleção dos cromossomos para cruzamento foi realizada através de um torneio simples;
- dois cromossomos selecionados possuem a probabilidade de cruzamento de 90%;
- para o cruzamento dos cromossomos, será utilizada a estratégia de dois pontos;
- a mutação dos cromossomos utilizará de uma estratégia aleatória;

Todas as demais diretivas aplicadas ao AG terão seus valores ideais verificados através de um procedimento de calibragem que investigará os seguintes parâmetros:

- O tamanho da população do AG;

- o número de gerações;
- a taxa de mutação aplicada;
- o modo de sobrevivência ou reprodução utilizada.

Dentre os métodos de sobrevivência considerados para o ajuste dos parâmetros serão considerados modos de sobrevivência clássicos como o *geracional*, *elitista* e o *steady-state*. Será investigado ainda um modo alternativo que mantém uma população principal e outra intermediária; esta é alimentada a cada geração pelas operações de seleção, cruzamento, mutação e armazena os filhos; aquela que mantém os melhores indivíduos que vieram da geração anterior sendo responsável pela carga genética para as novas gerações e armazena os pais. A alteração na população principal é dada pela competição entre pais e filhos que estão na mesma posição de suas respectivas populações. Tal característica se assemelha à estratégia utilizada no modo de sobrevivência *steady-state*, que também realiza a competição entre pais e filhos. A principal diferença é que nesta metodologia a competição de cada pai ocorre com apenas um filho. Para facilitar a retomada desse conceito, toda referência a essa estratégia de sobrevivência será realizada através do termo *posicional*.

3.5 Metodologia experimental

Os experimentos aplicados à metodologia consideram duas diferentes abordagens. Uma primeira que, através de experimentos epidêmicos cuja a origem da contaminação é conhecida, auxilia na determinação dos melhores parâmetros a serem aplicados ao AG. Já na segunda abordagem, que utiliza de experimentos cuja a origem infecciosa é desconhecida, concluem sobre os melhores parâmetros e averigua sobre a estabilidade da metodologia.

Os parâmetros considerados e seus possíveis valores a serem utilizados no AG estão dispostos na lista abaixo.

- População do AG: $p_{ag} = \{50, 100, 200\}$;
- gerações do AG: $g_{ag} = \{50, 100, 200\}$;
- taxa de mutação: $m_{ag} = \{0.05, 0.1\}$;
- modos de sobrevivência do AG: *geracional*, *elitista*, *posicional* e *steady-state*;
- *gap* no modo “*steady-state*”: $s_{ag} = \{0.0, 0.1, \dots, 0.9\}$.

Inicialmente os parâmetros são verificados em uma bateria de experimentos que utilizam o *experimento base* para avaliação. Para cada conjunto de parâmetros, infecta-se

vértice 21 da rede *moreno highschool* e executa-se a metodologia 100 vezes. O sucesso é registrado se o AG propor os oito vizinhos diretos do vértice como solução. A taxa de sucesso é definido em percentual.

Nesta fase dos experimentos, decidiu-se pela utilização de duas estratégias diferentes para a sumarização dos resultados. Uma que considera a premissa da metodologia, onde a frequência dos vértices contidos nos 10% melhores cromossomos da última geração é considerada. Outra que considera apenas a melhor solução obtida na última geração. O objetivo do comparativo é constatar se a estratégia de considerar parte dos melhores cromossomos é exitosa e acertada.

Os melhores conjuntos de parâmetros observados na bateria de testes anterior serão escolhidos para uma série mais detalhada de testes. Averigua-se a competência do conjunto de parâmetros e do algoritmo de otimização em obter a *contenção direta* de diversos nós. Assim, todos os vértices das redes *moreno highschool* e *moreno innovation* que possuírem grau superior a dois serão considerados. Testou-se respectivamente 67 e 230 nós dessas redes.

Bons resultados nos experimentos até então propostos não garantem sucesso da metodologia, visto que seu principal objetivo é propor vacinados para uma epidemia de origem desconhecida. É necessário que o modelo seja hábil na busca dos indivíduos de maior impacto na rede e que causem efeitos na disseminação da doença.

O princípio de se buscar diferentes perspectivas da dinâmica de espalhamento na rede é obtido através das múltiplas execuções do AG com diferentes grupos de infectados. Neste trabalho não será realizado um estudo sobre os impactos da mudança de tamanho do grupo de infectados, mantendo-se fixo a porcentagem de infecção em 5% da população. Assumiu-se serem necessárias 100 execuções do AG.

Resumidamente, análogo a um processo de Monte Carlo, o AG é executado 100 vezes. A cada vez infecta-se um grupo de 5% dos indivíduos aleatoriamente. Os vértices mais frequentes contidos nos cromossomos extraídos das execuções do AG são selecionados até o limite de doses disponíveis o que determina o conjunto de vacinados para uma epidemia de *origem desconhecida*.

Como apresentado na Seção 3.1, o conceito de estabilidade investiga a capacidade do modelo em prover soluções semelhantes para diferentes execuções. A *estabilidade vertical* será verificada através de experimentos na rede *moreno highschool* em busca da solução de uma epidemia de *origem desconhecida*. Serão realizadas 100 execuções com os conjuntos de parâmetros selecionados para uma vacinação de 15% da população da rede. É desejável que haja homogeneidade na escolha dos indivíduos, caso contrário qualquer resultado obtido poderia ser devido ao acaso.

Por fim, a *estabilidade horizontal* verifica se os indivíduos propostos pelo mo-

delo persistem com o incremento do percentual de vacinados. É desejável que os indivíduos indicados em uma vacinação de 15% da população também sejam escolhidos com 20% de vacinação. Essa característica será verificada através de testes nas redes *moreno highschool* e *moreno innovation* e com a variação do percentual de vacinados em $\{0.15, 0.2, 0.25, 0.3, 0.35, 0.4\}$. A quantidade de vacinas disponíveis em função do tamanho da rede pode ser observada na Tabela 3. As relações entre diferentes taxas de vacinação serão representadas através de diagramas de Venn e discutidas com base na intersecção dos indivíduos desses grupos propostos.

Tabela 3 – Quantidade de vacinas disponíveis por rede

Rede	15%	20%	25%	30%	35%	40%
moreno highschool	10	14	17	21	24	28
moreno innovation	36	48	60	72	84	96

3.6 Resultados

3.6.1 Experimentos que envolvem uma epidemia de origem conhecida

Os resultados da primeira bateria de testes para avaliação dos parâmetros do AG podem ser vistos na Tabela 10 do Apêndice A. Os resultados obtidos foram registrados como a razão de casos de sucesso pelo número de experimentos realizados e fixados em 100.

Nesta bateria de testes foram consideradas as duas estratégias de sumarização dos vértices para uma execução do AG. Ficou evidente que a estratégia que considera a frequência dos indivíduos que pertencem aos 10% melhores cromossomos da última geração é acertada e superior ao melhor cromossomo obtido na execução. Enquanto em alguns momentos se observou uma taxa de sucesso de 100% com a utilização dos 10% melhores, a taxa máxima obtida com o melhor cromossomo foi de 41%.

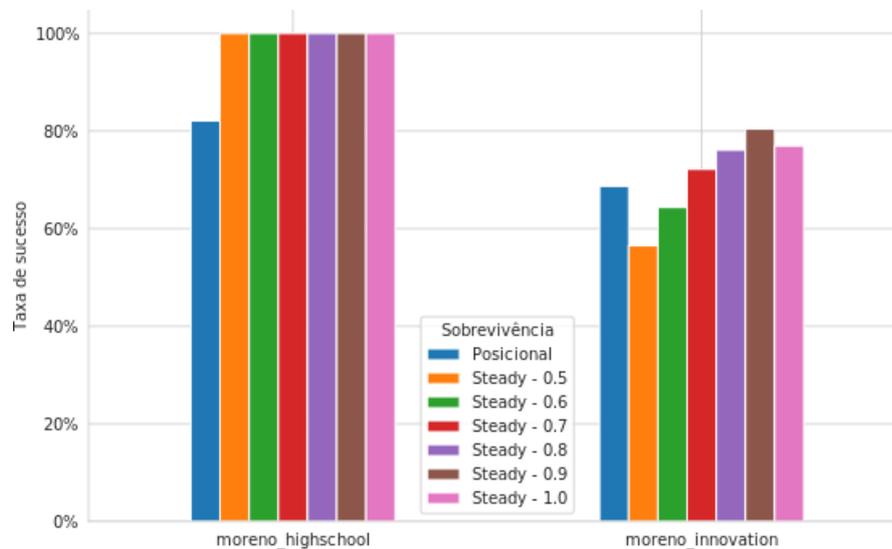
Diversos conjuntos de parâmetros resultaram em uma taxa de 100% de acerto. Viu-se até então a predominância do modo de sobrevivência *steady-state* entre os melhores conjuntos de parâmetros. Seus resultados estão presentes em 18 dos 20 melhores agrupamentos. A estratégia *posicional* apareceu com bons resultados nas outras duas posições. Esses grupos possuem em comum os parâmetros $p_{ag} = 200$, $g_{ag} = 200$ e $m_{ag} = 0.05$.

Todos os bons resultados do modo *steady-state* ocorreram com um alto percentual de *gap*, que representa o número de cromossomos novos gerados para concorrerem à sobrevivência. Altos valores desse parâmetro indicam maior tempo computacional devido ao maior número de avaliações de cromossomos e ao custo de ordenação da população de

soluções necessários para o avanço de cada geração. O ordenação decorre da necessidade de se selecionar os melhores cromossomos dentre todos os pais e filhos.

Os grupos de parâmetros que apresentaram taxa de sucesso igual a 100% foram selecionados para avaliação na próxima bateria de testes que mensura a capacidade do algoritmo em prover solução para a contenção direta dos demais vértices das redes *moreno highschool* e *moreno innovation* que possuam grau maior que dois.

Os resultados da segunda fase de experimentos para a *contenção direta* podem ser vistos na Figura 11. Os resultados são discutidos considerando que a solução é provida a partir dos vértices contidos nos 10% melhores cromossomos da execução do AG. A taxa de sucesso da contenção direta dos vértices da rede *moreno highschool* é bem próxima ou igual a 100% em todos os conjuntos de parâmetros avaliados. Já na rede *moreno innovation* a maior taxa de sucesso foi 80%.



Fonte: O autor

Figura 11 – Taxa de sucesso dos experimentos de contenção direta estendidos e agregados pelo *histograma 10%*

Viu-se novamente que os conjuntos que utilizam a sobrevivência *steady-state* apresentaram as maiores taxas de sucesso. Os melhores resultados foram vistos para o grupo que utilizou o *gap* de 0.9, que apresentou taxas de 100% e 80% para as redes *moreno highschool* e *moreno innovation*. O grupo de parâmetros que utiliza da sobrevivência *posicional* apresentou taxa de sucesso entre as redes com um intervalo menos discordante, com 80% na rede *moreno highschool* e 70% na rede *moreno innovation*.

Dado esse aparente equilíbrio de resultados do conjunto de parâmetros que utiliza o modo *posicional* junto a frequência e as maiores taxas absolutas de sucesso do modo *steady-state*, optou-se pela seleção dos conjuntos de parâmetros descritos na Tabela 4 para a execução de experimentos que analisassem a estabilidade da metodologia.

Tabela 4 – Conjunto de parâmetros selecionados para simulação de epidemias de *origem desconhecida* e análise de estabilidade na escolha de indivíduos. Os parâmetros listados têm em comum os valores $p_{ag} = 200$, $g_{ag} = 200$ e $m_{ag} = 0.05$

Sobrevivência	Gap
Posicional	-
Steady-state	1.0
Steady-state	0.9
Steady-state	0.8
Steady-state	0.7

Espera-se que a bateria de experimentos seguinte possa auxiliar na definição do melhor conjunto a ser considerado.

3.6.2 Experimentos que envolvem uma epidemia de origem desconhecida

A partir dos resultados obtidos nos experimentos da Seção 3.6.1 foram escolhidos os conjuntos de parâmetros listados na Tabela 4. Tais conjuntos apresentaram bons resultados em experimentos que buscavam solução para epidemias de origem conhecida e que por isso utilizavam de uma única execução do AG.

Nesta seção, como a origem da epidemia é desconhecida, necessita-se de diferentes perspectivas da disseminação da doença. Com isso, assume-se 100 execuções do AG, infectando-se grupos aleatórios de 5% dos vértices a cada execução. Os 10% melhores cromossomos de cada execução são extraídos e o histograma para avaliação dos vértices mais frequentes define o grupo de vacinados, selecionados até o limite de doses disponíveis.

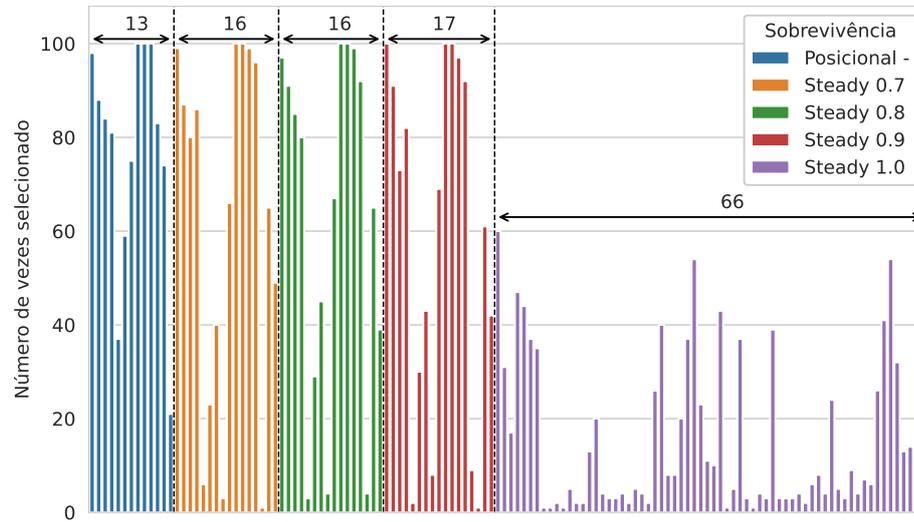
O primeiro experimento considerado para avaliação da *estabilidade vertical* é dado pela realização de 100 simulações da metodologia GASIR que tem como objetivo prover uma cobertura vacinal de 15%, ou seja, dez indivíduos da rede *moreno highschool*. É desejável que a metodologia retorne conjuntos semelhantes de indivíduos como solução de cada simulação.

A Figura 12 apresenta a frequência da seleção dos vértices como parte da solução. Vê-se por exemplo que nos testes realizados com o modo de sobrevivência *posicional*, dez indivíduos foram selecionados em mais de 70% das simulações. Os testes indicaram ainda que foram escolhidos apenas 13 indivíduos distintos dentre todas as execuções.

Os resultados do modo *steady-state* são bastante equivalentes quando se compara os parâmetros $gap = \{0.7, 0.8, 0.9\}$. Todos mantiveram 8 indivíduos selecionados em mais de 70% das simulações e escolheram entre 16 e 17 indivíduos distintos.

Como as soluções são fruto de um algoritmo de otimização, onde não existe a garantia de otimalidade, é esperado resultados diferentes entre as execuções. A melhora da intensificação dos parâmetros é observada quando os resultados acima descritos são

comparados ao conjunto que utiliza o $gap = 1.0$. Nesse caso, o vértice que se manteve mais frequente entre as soluções foi observado em 60% das simulações. Observou-se ainda a seleção de 66 diferentes vértices como parte de alguma das soluções, fator que constata a importância desse tipo de análise. Esse conjunto de parâmetros apresentou uma diversificação de resultados elevada, característica indesejável ao problema proposto.



Fonte: O autor

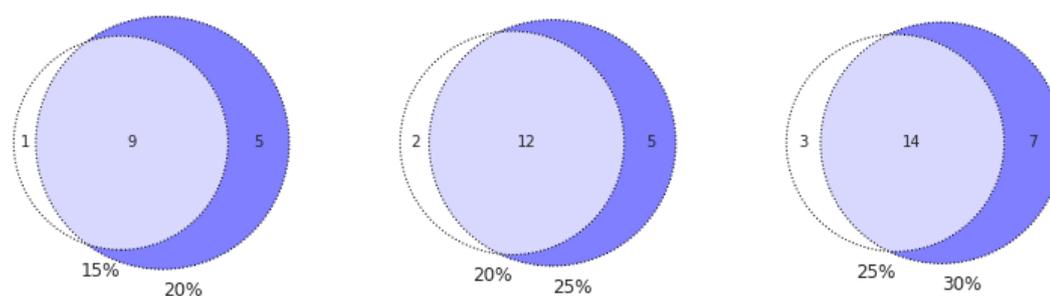
Figura 12 – Frequência dos indivíduos selecionados na análise vertical. As contagem entre os intervalos destacados registram o número de vértices selecionados dentre as 100 execuções do modelo.

Os conjuntos de parâmetros que apresentaram um menor número de indivíduos distintos selecionados foram classificados como satisfatórios. Dada a consistência levemente superior do modo de sobrevivência *posicional*, optou-se por adotá-lo como parte do melhor conjunto de parâmetros a ser utilizado na metodologia GASIR.

Tabela 5 – Parâmetros do AG aplicados a todos os experimentos do novo modelo GASIR.

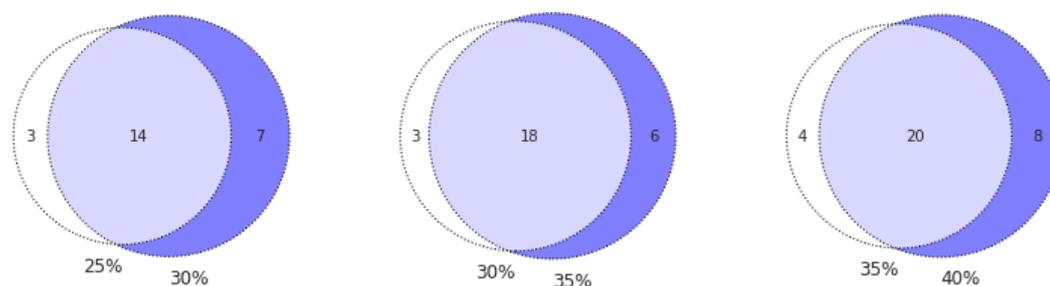
Parâmetro	Valor
p_{ag}	200
g_{ag}	200
m_{ag}	0.05
s_{ag}	-
Taxa de Cruzamento	0.9
Tipo de Cruzamento	2 pontos
Modo de sobrevivência	Posicional

Os experimentos que averiguam a estabilidade horizontal do modelo GASIR utilizaram como parâmetros os dados da Tabela 5 e as demais diretivas previamente definidas para epidemias de *origem desconhecida*. Executou-se a metodologia para as redes *moreno highschool* e *moreno innovation* variando-se o percentual de vacinação conforme descrito na Tabela 3.



Fonte: O autor

Figura 13 – Intersecção de grupos de 15%, 20%, 25% e 30% de vacinados na rede *moreno highschool*



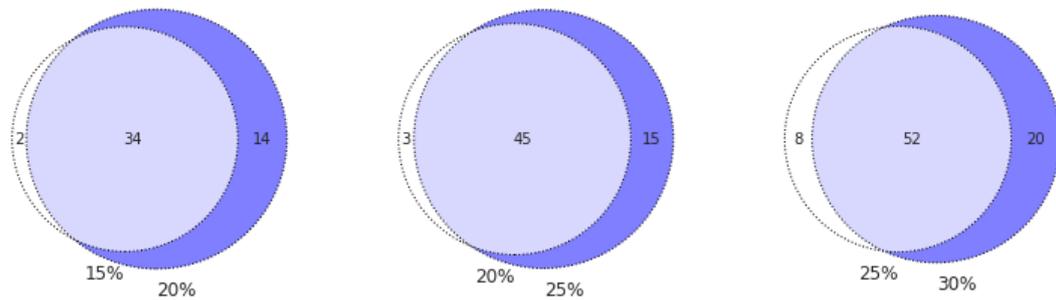
Fonte: O autor

Figura 14 – Intersecção de grupos de 25%, 30%, 35% e 40% de vacinados na rede *moreno highschool*

Os resultados obtidos para a rede *moreno highschool* podem ser vistos nas Figuras 13 e 14. Assim como na estabilidade vertical, a estabilidade não é absoluta e alguns indivíduos selecionados em campanhas menores não persistiram na campanha maior. Viu-se alguma variação desses indivíduos “esquecidos”. Aproximadamente 85.4% dos indivíduos propostos em uma campanha foram mantidos com o aumento do percentual de vacinados, com um desvio padrão de 2%.

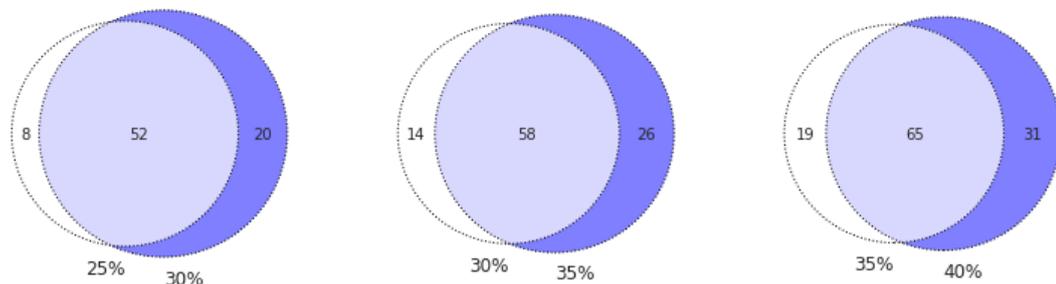
Cenário semelhante foi observado nos resultados da rede *moreno innovation* e pode ser visto nas Figuras 15 e 16. O comportamento do modelo de não selecionar alguns indivíduos indicados na campanha anterior persiste a partir do primeiro comparativo. Contudo, os valores médios de permanência foram próximos daqueles vistos na rede anterior. Foi observada a média de persistência dos vértices de aproximadamente 87%, com desvio padrão de 7%.

A possibilidade de execução do AG para um percentual maior de vacinados para depois se selecionar um número menor de vértices foi uma alternativa considerada para a metodologia. Presumiu-se que o objetivo primário da metodologia, que ajusta a seleção do vértices para um número restrito de vacinas “*in silico*”, é o seu principal diferencial. O maior número de indivíduos suscetíveis faz com que as medições da função objetivo possuam escala maior, o que cria um cenário mais favorável para o AG, além de permitir uma melhor percepção do grau de estabilidade da metodologia. A Figura 17 ilustra um



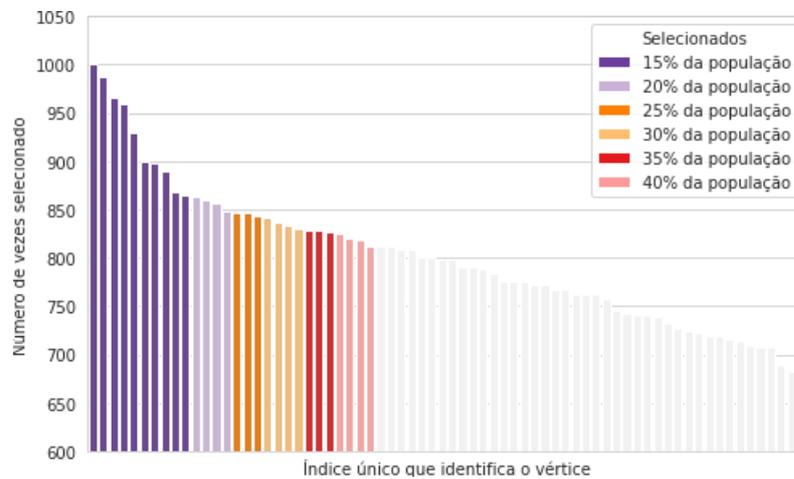
Fonte: O autor

Figura 15 – Interseção de grupos de 15%, 20%, 25% e 30% de vacinados na rede *moreno innovation*



Fonte: O autor

Figura 16 – Interseção de grupos de 25%, 30%, 35% e 40% de vacinados na rede *moreno innovation*



Fonte: O autor

Figura 17 – Histograma final do GASIR para a vacinação de 40% da população na rede *moreno highschool*. A seleção deve ser considerada complementar, ou seja, os indivíduos destacados em 15% da população estão contidos nos demais percentuais destacados.

histograma de frequência dos vértices contidos nos 10% melhores cromossomos de cada uma das 100 execuções do AG para a vacinação de 40% na rede *moreno highschool*. Diferentes cores indicam os indivíduos que seriam selecionados conforme o percentual de vacina disponível.

3.7 Considerações finais do capítulo

Este capítulo descreveu todas as nuances da metodologia que propõe uma escolha sistematizada de vértices a serem imunizados em uma rede complexa a partir dos parâmetros da doença e ajustados ao limite de doses disponíveis. O problema foi modelado para se utilizar de um algoritmo genético como ferramenta para a proposição de diversas combinações de vacinados.

O modelo utiliza de uma função de avaliação estocástica, característica que dificulta o trabalho do algoritmo de otimização. Para contornar tal limitação, foram elaborados mecanismos para a reparação da avaliação das soluções, que busca considerar a pior avaliação possível para um problema cujo alvo é minimizar a função objetivo.

A metodologia considera ainda múltiplas execuções do mecanismo de otimização para a proposição da imunização em uma rede independentemente, da origem da doença. A análise da frequência dos vértices propostos como solução, através de um histograma, define os melhores vértices para a vacinação.

Um estudo dos parâmetros do algoritmo de otimização foi realizado. Os valores selecionados atendem duas premissas: são valores que apresentaram resultados esperados em um cenário onde a solução ótima é conhecida e que sinalizaram trazer maior estabilidade para os resultados propostos pela metodologia.

Ficou evidente que a utilização do histograma como ferramenta auxiliar no processo de seleção dos melhores indivíduos é uma decisão acertada e que facilita o trabalho do algoritmo de otimização.

Os níveis de estabilidade indicados com os experimentos não são absolutos, mas demonstram o bom funcionamento da metodologia. Com os atuais ajustes propostos, tal estabilidade poderia ficar mais destacada com um número maior de execuções do AG, mas essa opção não foi considerada devido ao amplo tempo computacional envolvido. Uma única execução da metodologia na rede *moreno highschool* para prover uma lista de vacinados igual a 15% da população é executada em aproximadamente 7400 segundos.

Optou-se por prosseguir com a análise dos impactos que a vacinação dos indivíduos propostos pela metodologia causa em epidemias simuladas. Tal efetividade será comparada a outros critérios de vacinação. Caso a solução do GASIR se apresente superior, é desejável a aplicação de melhorias que busquem a redução do tempo computacional e até a melhora dos níveis de estabilidade das soluções propostas.

4 Efetividade dos critérios de vacinação

Este Capítulo propõe e discute uma série de experimentos que anseia avaliar a efetividade do GASIR e de outros critérios de vacinação na escolha de indivíduos ideais a serem vacinados em uma rede complexa. Serão considerados critérios que utilizem de medidas de centralidade e de aleatoriedade para a escolha de indivíduos. A ausência de vacinação também é utilizada para ilustrar a importância da vacinação.

O conjunto de experimentos descritos e discutidos buscará resultados mais contundentes sobre a efetividade desses critérios sobre diferentes perspectivas do processo epidêmico simulado nas redes selecionadas.

4.1 A metodologia dos experimentos

4.1.1 Descrição das redes avaliadas

Para que se tenha um melhor entendimento dos efeitos da escolha de indivíduos providos pelos critérios analisados, optou-se por estender o número de redes experimentadas. Decidiu-se pelo uso de redes reais e de redes artificiais heterogêneas construídas através do modelo *Barabási-Albert* (BARABASI; ALBERT, 1999). Algumas dessas redes são originalmente dirigidas, podendo existir arestas paralelas com sentidos opostos. Nesses casos, a direção e uma das arestas paralelas serão removidas.

Foram escolhidas quatro redes reais clássicas no estudo de redes complexas e descritas a seguir. O detalhamento da construção e o tipo de interações que essas representam foram obtidos nos trabalhos que as introduzem e no repositório (PEIXOTO, 2020) que armazena informações de diversas redes complexas clássicas.

Rede moreno highschool: Rede direcionada que registra a amizade entre meninos de uma escola secundária em *Illinois*. A rede foi formada questionando a cada um dos rapazes quem eram seus amigos. Foi inicialmente publicada e estudada no trabalho de (COLEMAN, 1964).

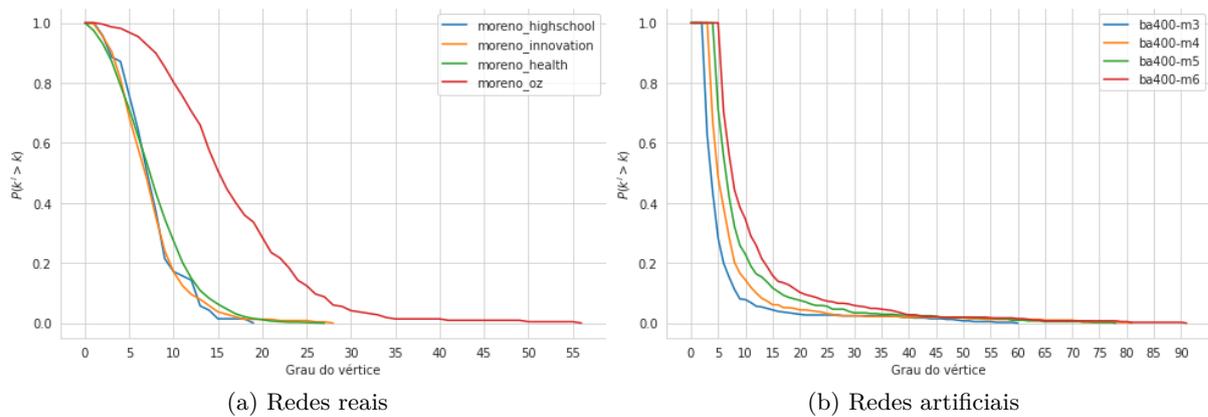
Rede moreno innovation: Rede também direcionada que registrou as relações de confiança entre físicos que residiam em quatro cidades do meio-oeste dos Estados Unidos. As arestas registram que um físico i confia e pede conselhos a outro j . Foi formalmente citado em (COLEMAN et al., 1957)

Rede moreno oz: Rede dirigida de amizades entre estudantes que moram em uma residência universitária na *Australian National University*. A direção da aresta define

que um estudante i declarou outro j como amigo. É fruto do trabalho de (FREEMAN et al., 1998).

Rede moreno health: Rede também direcionada obtida através uma pesquisa social com alunos do ensino médio em 1994 e que representam 84 diferentes comunidades. Esse extenso trabalho de pesquisa foi estudado e publicado em (MOODY, 2001).

A principal motivação para a escolha de redes artificiais para composição dos experimentos é a de se trazer uma maior diversidade de ambiente através de um modelo capaz de produzir redes heterogêneas e com características comparáveis às reais. A Figura 18 ilustra as semelhanças do comportamento da curva que representa a função de distribuição acumulada complementar (CCDF) e que remete ao grau dos vértices das redes consideradas.



Fonte: O autor

Figura 18 – Distribuição de grau complementar acumulada das redes estudas

Foram criadas quatro redes segundo o modelo de *Barabási-Albert* que adicionavam m arestas para cada um dos 400 vértices adicionados, tal que $m \in \{3, 4, 5, 6\}$. Para que se evite qualquer enviesamento na anexação preferencial do modelo, cada uma das redes foi iniciada com um grafo semente completo de $m_0 = 10$ vértices. Essa escolha foi inspirada nos argumentos dispostos no trabalho de (HELD et al., 2014).

A Tabela 6 mostra os parâmetros estruturais e topológicos das redes utilizadas nesse trabalho.

É importante notar que o coeficiente de assortatividade das redes escolhidas é muito próximo de zero. O trabalho de (NEWMAN, 2002) descreve que redes assortativas são bastante robustas enquanto as redes disassortativas possuem baixa robustez. A escolha de redes com valores que tendem a neutro é intencional para que se crie um cenário mais equilibrado para os experimentos, sem dificultar ou facilitar o trabalho do algoritmo de otimização utilizado no GASIR.

Tabela 6 – Dados gerais das redes utilizadas para simulações. Considere que as redes possuem $|V|$ vértices, $|E|$ arestas, \bar{k} grau médio, k_{max} de grau máximo e k_{min} de grau mínimo. As colunas c e r representam o coeficiente de *clustering* e assortatividade respectivamente.

Rede	$ V $	$ E $	\bar{k}	k_{max}	k_{min}	c	r
moreno highschool	70	274	7.83	19	2	0.40	0.01
moreno innovation	241	929	7.71	28	1	0.25	-0.01
moreno health	2539	10455	8.24	27	1	0.14	0.01
moreno oz	217	1839	16.95	56	2	0.30	0.00
ba400-m3	410	1245	6.07	60	3	0.06	-0.06
ba400-m4	410	1645	8.02	81	4	0.07	-0.05
ba400-m5	410	2045	9.98	78	5	0.07	-0.03
ba400-m6	410	2445	11.93	91	6	0.08	-0.03

4.1.2 Metodologia aplicada para os experimentos

A avaliação da efetividade dos critérios de vacinação será realizada através de uma análise comparativa. Todas as simulações serão realizadas através do modelo SIR que provê parâmetros para que se defina uma probabilidade de infecção distinta para cada aresta entre dois vértices quaisquer i e j da rede, dada pelo parâmetro β_{ij} ; a probabilidade de recuperação distinta para cada vértice i , dada por γ_i ; e a probabilidade de infecção espontânea de cada vértice, dada por h_i . Todos os experimentos aplicados basearam-se em parâmetros epidêmicos estimados para uma epidemia de *Influenza* e obtidos através de informações dos trabalhos de Bresee et al. (2011) e Coburn et al. (2009). Assim, a taxa de infecção β assumida foi de 0.2857 em todas as arestas das redes, a taxa de recuperação γ de 0.1428 para todos os vértices e a atribuição da taxa de infecção espontânea $h_i = 0$.

Serão experimentados diferentes percentuais de indivíduos vacinados, ou pv , em cada uma das redes com valores entre 15% e 40% e que variam de 5% em 5%. O percentual de indivíduos infectados iniciais será sempre de 5%. A Tabela 7 indica o número de indivíduos infectados e vacinados por rede e por percentual de imunização.

Os critérios de vacinação considerados para este estudo comparativo são listados a seguir. Em sua maioria, serão utilizadas medidas de centralidade para a escolha dos indivíduos a serem vacinados. Essas medidas são funções que definem numericamente a importância de um vértice na rede ou do grupo a que ele pertence. A descrição dos critérios de vacinação considerados é dada a seguir.

- Vértices que possuem a maior centralidade de *grau* (*degree*) ou \mathcal{C}_{pv}^D ;
- vértices que possuem a maior centralidade de *betweenness* ou \mathcal{C}_{pv}^B ;
- vértices que possuem a maior centralidade de *pagerank* ou \mathcal{C}_{pv}^P ;
- vértices que possuem a maior centralidade de autovetor (*eigenvector*) ou \mathcal{C}_{pv}^E ;

- vértices propostos pela metodologia GASIR ou \mathcal{C}_{pv}^G ;
- vacinação randômica ou \mathcal{C}_{pv}^R ;
- ausência de vacinação ou \mathcal{C}^\emptyset .

A primeira característica a ser averiguada trata da similaridade entre os indivíduos selecionados por critério de vacinação. Cada critério formará uma lista de indivíduos a serem vacinados e que têm sua cardinalidade restrita ao número de vacinas disponíveis. Torna-se necessário comparar aos pares listas parciais e que podem ser inclusive disjuntas, características que limitam o uso de técnicas de correlação estatística.

Para uma análise eficaz dada essas restrições, optou-se por utilizar o RBO, ou *Rank-Biased Overlap*. Tal medida foi proposta por (WEBBER et al., 2010) para atuar com classificações ordinais e indefinidas. Por se tratar de uma medida baseada em ordenação e intersecção, acredita-se que este método seja mais adequado para a coleta das impressões sobre a formação da lista de indivíduos de cada critério e de sua ordenação. A medida respeita um intervalo numérico entre zero e um onde maiores valores implicam em maior nível de similaridade entre as listas. O nível de similaridade dado pelo RBO será avaliado junto à intersecção dos indivíduos entre os critérios avaliados.

Como a natureza estocástica do GASIR pode resultar em soluções distintas a cada percentual de vacinação considerado, definiu-se ser mais adequada a comparação entre conjuntos de vacinados propostos para o mesmo percentual de vacinação.

Tabela 7 – Redes analisadas: infectados e vacinas

Rede	Infectados	Vacinados					
		15%	20%	25%	30%	35%	40%
moreno highschool	3	10	14	17	21	24	28
moreno innovation	12	36	48	60	72	84	96
moreno health	126	380	507	634	761	888	1015
moreno oz	10	32	43	54	65	75	86
ba400-m3	20	61	82	102	123	143	164
ba400-m4	20	61	82	102	123	143	164
ba400-m5	20	61	82	102	123	143	164
ba400-m6	20	61	82	102	123	143	164

O segundo aspecto a ser averiguado trata da eficácia e o impacto que os indivíduos escolhidos pelos critérios trazem ao processo de disseminação da doença na rede. Como foi visto até o momento, o desconhecimento da origem da epidemia e o comportamento estocástico do SIR dificultam em muito a análise da vacinação. Uma única simulação do processo epidêmico pode gerar valores distintos e pouco conclusivos. Com isso, estipulou-se uma série de passos a serem tomados para a avaliação dos diferentes critérios de vacinação.

Considere que cada uma das redes é representada por $G = (V, E)$, onde $n = |V|$ representa o número de vértices da rede e V o conjunto de todos os vértices. Os passos a seguir apresentam a dinâmica dos experimentos realizados para cada percentual de vacinados pv e que serão posteriormente analisados:

1. Forma-se o grupo de indivíduos candidatos a serem infectados X constituído pela diferença de todos os indivíduos da rede V e os indivíduos selecionados por qualquer critério de vacinação, ou seja, $X = V - (\mathcal{C}_{pv}^D \cup \mathcal{C}_{pv}^B \cup \mathcal{C}_{pv}^P \cup \mathcal{C}_{pv}^E \cup \mathcal{C}_{pv}^G \cup \mathcal{C}_{pv}^R)$.
2. Um grupo de infectados Z_i é formado pela combinação de indivíduos contidos em X e que terá a cardinalidade correspondente a 5% do tamanho da rede, ou seja, igual a $0.05 * |V|$.
3. O conjunto de todos os grupos de infectados é constituído de 1000 grupos e dado por $Z = \{Z_i : i = 1, 2, 3, \dots, 1000\}$ tal que $Z_i \subset X$.
4. O SIR é utilizado para simular 100 vezes uma epidemia, utilizando-se de 60 iterações síncronas (passos) do modelo e aplicado para cada grupo de infectados Z_i e cada critério de vacinação da forma \mathcal{C}_{pv} .
5. A vacinação é aplicada nos vértices selecionados pelo critério antes da simulação com o SIR.
6. Será considerada apenas a simulação que atingiu o maior número de indivíduos da rede, o que registraria o pior caso possível dentre as simulações. Tal diretiva objetiva eliminar simulações que terminem precocemente.
7. Serão extraídas dessa simulação algumas variáveis dependentes que serão melhor descritas a seguir.

É importante destacar que, para o mesmo percentual de vacinação, os grupos de infectados permanecem os mesmos e são testados para cada critério de vacinação discutido. Portanto, novos percentuais de vacinação gerarão novos grupos de infectados.

O pseudocódigo do método que reproduz todas as simulações necessárias para a avaliação da efetividade de cada um dos critérios é apresentado no Algoritmo 1. Vê-se que os valores *pico*, *afetados*, *beta* são salvos. Essas variáveis remetem às variáveis dependentes que serão analisadas e que significam, respectivamente:

- O pico (número máximo de todos os instantes) de indivíduos infectados;
- a soma de indivíduos afetados (recuperados ou infectados) e
- o β inferido da curva de infecção.

Cabe salientar que na contagem dos indivíduos afetados, os indivíduos selecionados para vacinação são contabilizados, por estarem no compartimento de recuperados.

Algoritmo 1: Algoritmo que representa a dinâmica das simulações de efetividade

```

/* itera sobre os diferentes percentuais de vacinados */
1 para cada  $pv \in \{0.15, 0.2, 0.25, 0.3, 0.35, 0.4\}$  faça
2    $X \leftarrow V - (\mathcal{C}_{pv}^D \cup \mathcal{C}_{pv}^B \cup \mathcal{C}_{pv}^P \cup \mathcal{C}_{pv}^E \cup \mathcal{C}_{pv}^G \cup \mathcal{C}_{pv}^R)$ ;
3    $Z \leftarrow infectados(X, 0.05)$ ; /* retorna grupos de infectado de 5% dos indivíduos */
   /* itera sobre cada conjunto de infectados */
4   para cada  $z \in Z$  faça
   /* itera sobre cada conjunto de vacinados conforme o percentual de vacinados */
5     para cada  $c \in \{\mathcal{C}_{pv}^D \cup \mathcal{C}_{pv}^B \cup \mathcal{C}_{pv}^P \cup \mathcal{C}_{pv}^E \cup \mathcal{C}_{pv}^G \cup \mathcal{C}_{pv}^R \cup \mathcal{C}^\emptyset\}$  faça
6        $\beta, a_{afetados}, p_{ico} \leftarrow 0$ ;
7       para  $i = 1$  até 100 faça
8          $s_{beta}, s_{afetados}, s_{pico} \leftarrow epidemia(z, c, 60)$  /* Simula a epidemia */
9         se  $s_{afetados} \geq a_{afetados}$  então
10          |  $\beta, a_{afetados}, p_{ico} \leftarrow s_{beta}, s_{afetados}, s_{pico}$ ;
11          fim
12        fim
13         $salvar(pv, z, c, \beta, a_{afetados}, p_{ico})$ ; /* Salva a pior simulação */
14      fim
15    fim
16 fim

```

Dentre essas variáveis, a inferência do β demanda um maior detalhamento. Com a probabilidade de recuperação γ mantida, esta inferência é aplicada à curva que representa número de infectados por instante de tempo do experimento.

Através de regressão com a técnica de mínimos quadrados não lineares, cada simulação estocástica teve seus dados ajustados à equação determinística do modelo SIR, definida na Equação 2.8. A percepção numérica aproximada com a estimativa do β anseia captar o “achatamento” característico da curva de infectados. A Figura 19 apresenta os efeitos que a alteração dos valores de β geram em uma curva de infectados quando aplicadas em equações determinísticas.

Como as simulações de epidemia em redes complexas são estocásticas, espera-se que esse ajuste da curva a um modelo determinístico seja uma boa estratégia para que se compare os efeitos de cada critério de vacinação. Posteriormente verificou-se que o trabalho de (NDIAYE et al., 2020) também utilizou o mesmo recurso para a predição de casos confirmados de COVID-19. A Figura 20 apresenta uma curva ajustada aos dados

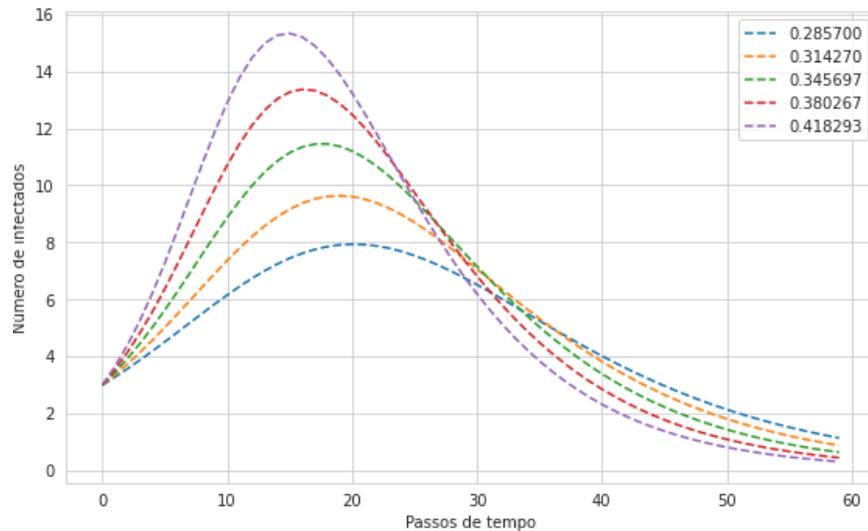


Figura 19 – Efeito da curva de infectados com a variação do β

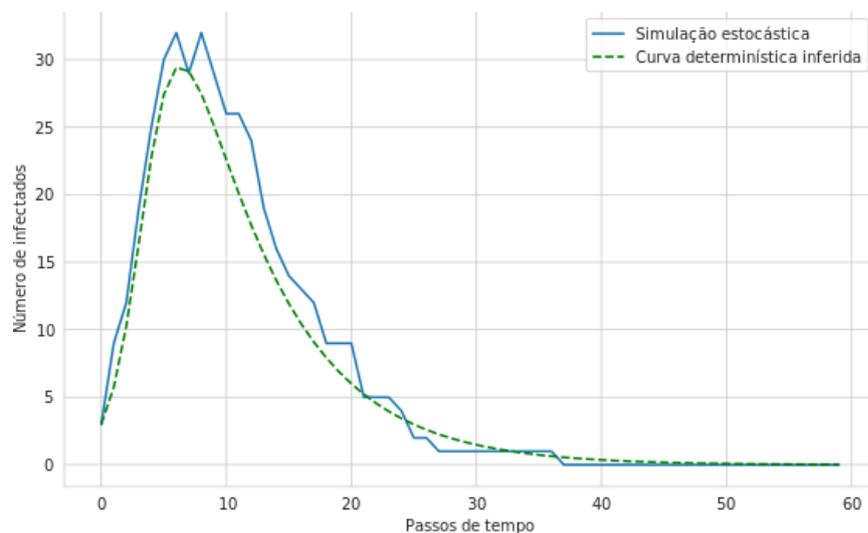


Figura 20 – Parâmetro β inferido em uma simulação na rede *moreno_highschool* com vacinação de 15% dos indivíduos. A curva em verde representa os dados ajustados. A simulação infectou os nós 3, 12 e 19.

de uma simulação com vacinação de 15% dos indivíduos na rede *moreno_highschool*.

Um dos principais pontos de preocupação deste trabalho foi sobre a perspectiva estatística que seria utilizada para a verificação das variáveis de saída estudadas. Como SIR aplicado é estocástico, é esperado uma alta variabilidade dos valores obtidos. A princípio, como no trabalho de [Hartvigsen et al. \(2007\)](#), imaginou-se que a averiguação da equivalência estatística dos experimentos baseada em média seria a melhor solução.

Das simulações aplicadas, extraiu-se três variáveis dependentes de 336 grupos diferentes de experimentos (oito redes, sete critérios de vacinação e seis diferentes percentuais de vacinação). Testes preliminares indicaram que, dessas 1008 amostras, em menos de 40% delas os dados das variáveis dependentes tinham aderência à normalidade. A avaliação foi realizada através do teste de *Shapiro-Wilk* ([SHAPIRO; WILK, 1965](#)), onde a rejeição da

hipótese nula H_0 indica que as amostras de dados não seguem uma distribuição normal.

Outro detalhe é que o princípio da homocedasticidade, que busca verificar a homogeneidade de variâncias entre as amostras agrupadas por rede e percentual de vacinação, não foi atendido. Tal comportamento foi observado mesmo quando se considerou apenas os critérios que utilizam a centralidade e o GASIR para a escolha dos indivíduos. Essa restrição limita a utilização de técnicas de análise de variância que considerem a média como medida comparativa.

Presumiu-se que, a natureza do modelo epidemilógico SIR, que possibilita o retorno de resultados aleatórios, seria a principal causa desse comportamento estatístico observado.

Foram realizados diversos procedimentos para o ajuste dos dados a uma distribuição normal. Aplicou-se a remoção dos *outliers* e algumas técnicas de transformação de dados como a transformação logarítmica e transformação de *Box-Cox* (BOX; COX, 1964) que apresentaram correções para um número muito pequeno de amostras. A utilização de transformação de dados para aderência à normalidade é discutida no trabalho de Oliveira et al. (2018) e um dos pontos considerados é sobre a mudança de escala que este procedimentos pode gerar.

Tendo em vista estas características observadas nas amostras, decidiu-se pela utilização da abordagem estatística não paramétrica para avaliação da equivalência entre os diferentes critérios de vacinação. Portanto, assume-se que a mediana e a amplitude interquartil serão as medidas consideradas na comparação das amostras.

Assim, optou-se por realizar o teste de Kruskal-Wallis (KRUSKAL; WALLIS, 1952) para análise de variância das medianas dos diferentes critérios para que se verifique a existência de dominância estatística de pelo menos um. A hipótese nula H_0 será aceita se o valor de significância for dado por $p > 0.05$, representando que as medianas de todos o grupos avaliados são estatisticamente equivalentes.

Testes de *post-hoc* com o método de Mann-Whitney (MANN; WHITNEY, 1947) serão aplicados para averiguar se, conferidos aos pares, alguns destes critérios não possuem diferença estatística significativa. O valor de significância adotado será também de $p > 0.05$, o que indica o aceite da hipótese nula H_0 e representa que as medianas são estatisticamente equivalentes.

Os resultados dos testes estatísticos e a visualização gráfica das variáveis permitirão a análise da efetividade dos critérios discutidos neste trabalho.

4.2 Resultados

4.2.1 Similaridades e intersecção

A similaridade e a intersecção entre os critérios de vacinação pode auxiliar na correta interpretação dos resultados experimentais, fruto das simulações epidêmicas. Por razões óbvias, esta análise não inclui o critério \mathcal{C}^θ devido a ausência de indivíduos vacinados.

As Figuras 28 a 43 presentes no Apêndice B ilustram, através de um mapa de calor, o nível de similaridade rbo e a intersecção dos indivíduos entre os critérios tomados aos pares.

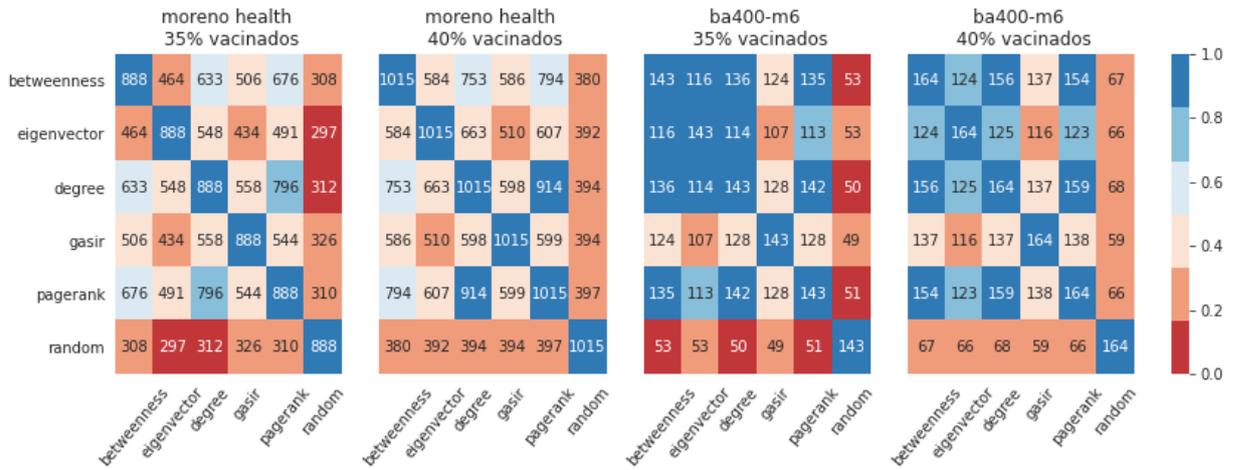
Foi observado que em todas as redes artificiais os níveis de similaridade entre \mathcal{C}_{pv}^G e todos os demais critérios foram inferiores a 0.5. Os níveis de intersecção são maiores e mais consistentes comparando-se aos critérios \mathcal{C}_{pv}^D , \mathcal{C}_{pv}^P e \mathcal{C}_{pv}^B e variam entre 65% e 92%. Notou-se ainda que, com o aumento do grau médio das redes, os critérios \mathcal{C}_{pv}^E e \mathcal{C}_{pv}^G aumentaram o número de elementos comuns com os demais critérios de centralidade.

Já na análise das redes reais, observou-se que os níveis de similaridade do critério \mathcal{C}_{pv}^G possuem a mesma tendência das redes artificiais, ou seja, $rbo < 0.6$. Os níveis de intersecção entre os diferentes critérios também variam bastante, e são normalmente maiores quando comparados a \mathcal{C}_{pv}^P e \mathcal{C}_{pv}^D . Neste escopo, o aumento da intersecção não ocorreu com o aumento do grau médio. Apenas nas redes *moreno oz* e *moreno health*, viu-se que os níveis de intersecção subiram com o aumento do percentual de vacinados.

A Figura 21 ilustra a diferença entre os níveis de similaridade registrados em uma rede real e outra artificial com dois diferentes percentuais de vacinação. Vê-se que nas redes artificiais os valores obtidos pela medida rbo tendem a 1 entre todos os critérios que utilizam a centralidade como método de escolha. Isso registra que, mesmo que os conjuntos não possuam a maior intersecção em alguns momentos, que a ordenação dos elementos possui certa semelhança. Nas redes reais, tal característica não é tão explícita, devido à complexidade estrutural maior das redes, seja com maior coeficiente de agrupamento, maior grau médio ou topologia com presença de comunidades.

É possível concluir que o critério \mathcal{C}_{pv}^G possui uma natureza própria na escolha dos indivíduos caracterizada pela diferença ordinal da escolha dos vértices. Tal interpretação é consequência da menor similaridade rbo quando comparada a outros critérios. Embora haja menor similaridade, espera-se que os resultados de epidemias simuladas sejam semelhantes aos dos critérios \mathcal{C}_{pv}^D e \mathcal{C}_{pv}^P , exceto nas redes reais *moreno oz* e *moreno health*, dado o número de indivíduos comuns propostos pelos critérios.

Conforme esperado, os níveis de intersecção e de similaridade do critério \mathcal{C}_{pv}^R são baixos quando observados junto a qualquer outro critério.



Fonte: O autor

Figura 21 – Similaridade e intersecção dos critérios de vacinação - Redes reais x artificiais

O critério C_{pv}^E é, dos critérios baseados em centralidade, o que mais difere dos demais de mesma natureza. Seus valores de intersecção são comparáveis aos critérios C_{pv}^P e C_{pv}^D apenas nas rede *moreno oz*, *ba400-m5* e *ba400-m6*.

Os critérios C_{pv}^P e C_{pv}^D são os que mais se assemelham quando comparados em todas as redes experimentadas. Espera-se que o desempenho desses critérios seja semelhante.

4.2.2 Análise estatística das variáveis dependentes

O teste de Kruskal-Wallis é aplicado sobre as amostras de 1000 elementos agrupadas por rede, percentual de vacinação pv e critérios de vacinação na forma C_{pv} de cada uma das três variáveis dependentes estudadas. Ele verifica a hipótese nula H_0 de que as medianas das amostras agrupadas possuem equivalência estatística para uma mesma variável dependente. Para esta análise foram removidos os critérios C^\emptyset e C_{pv}^R por estes existirem apenas para fins comparativos de quão é acertada a decisão de se aplicar algum critério sistemático para vacinação.

Obteve-se como resultado a rejeição da hipótese nula H_0 a um nível de significância de 0.05, com grau de liberdade 4 para todos os 40 agrupamentos (8 redes, 5 critérios) das três variáveis dependentes. Isso representa que existe diferença estatística entre os critérios de vacinação propostos baseados em centralidade e o GASIR. A checagem dessa hipótese sinaliza que pelos menos dois critérios de vacinação adotados apresentam diferença. Para avaliação dos critérios aos pares, aplica-se um teste de *post-hoc* de *Mann-Whitney*.

No Apêndice C vê-se uma série de *boxplots* dos experimentos que ilustram os números obtidos pelas variáveis dependentes. As Tabelas 11, 12, e 13 do Apêndice A listam os critérios que apresentaram menor mediana agrupados por variável dependente, rede e percentual de vacinação. Eventuais equivalências estatísticas são registradas com seus respectivos *p-valor* e o valor apurado no teste de *Mann-Whitney*.

Não foi observada uma relação uniforme entre as variáveis dependentes analisadas. Especialmente nas redes reais, viu-se momentos onde os resultados da variável *afetados* geraram altos valores de amplitude interquartil, deixando praticamente indefinido qual o melhor critério. As diferenças foram então observadas através das variáveis *pico* e *beta* que indicam o achatamento da curva e a diminuição da taxa de contágio.

O comportamento dos resultados nas redes artificiais sugere similaridade entre os critérios C_{pv}^P e C_{pv}^D , fato verificado pelas frequentes indicações de equivalência estatística nos testes de *post-hoc*. Outra evidência observada é que o número de indivíduos que se mantiveram suscetíveis nestas simulações decresceu com o aumento do grau médio das redes estudadas. A Figura 22 ilustra o percentual de indivíduos que não receberam vacina e que não se infectaram após as simulações epidêmicas. O maior valor deste percentual indica um menor número de indivíduos que não tiveram contato com a doença, o que é desejável em uma epidemia. Vê-se que, com o aumento do número de arestas nas redes, o percentual de indivíduos que não adoeceu diminuiu, característica observada em todos os percentuais de vacinação considerados. Tal fator indica que o número de relações entre os indivíduos é um importante aspecto a ser considerado e que dificulta a contenção do espalhamento de uma doença. Uma vacinação de 40% da população nas redes *ba400-m3* e *ba400-m6* mantiveram suscetíveis aproximadamente 50% e 34% da população, respectivamente.

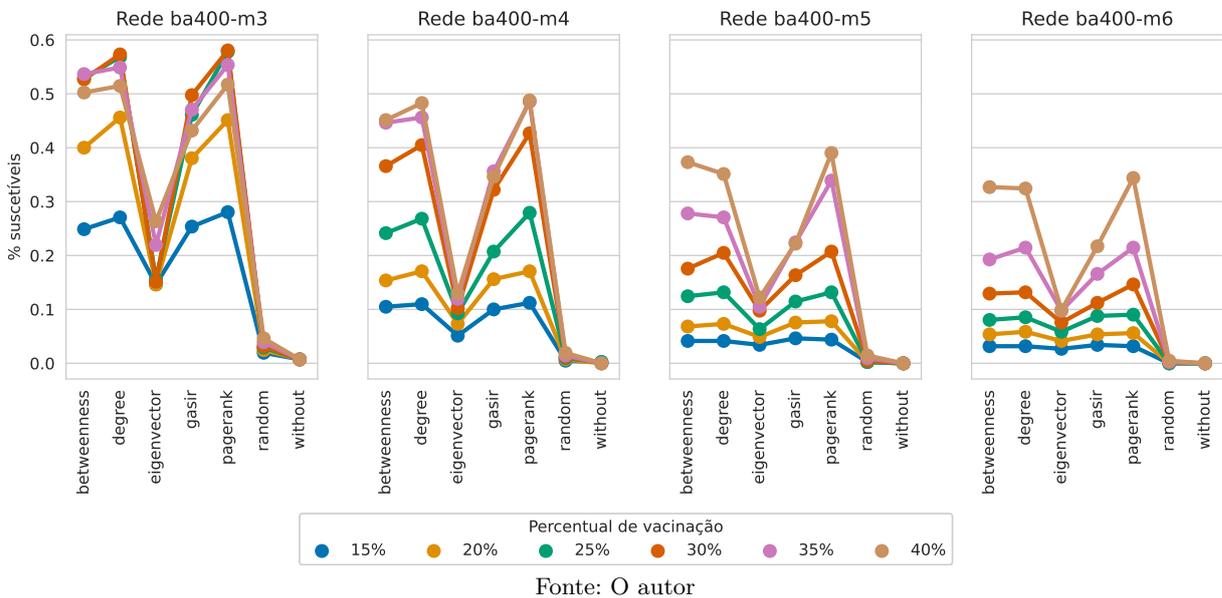
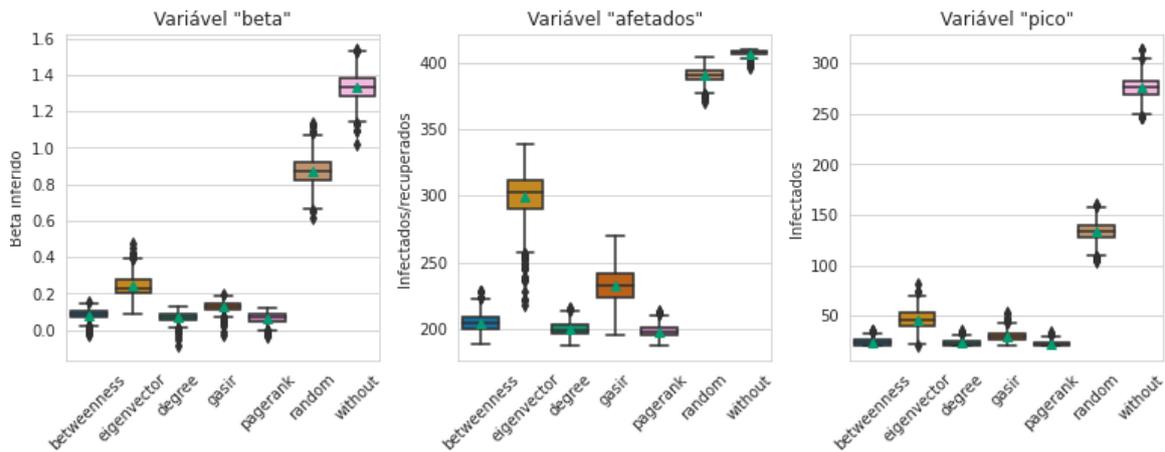


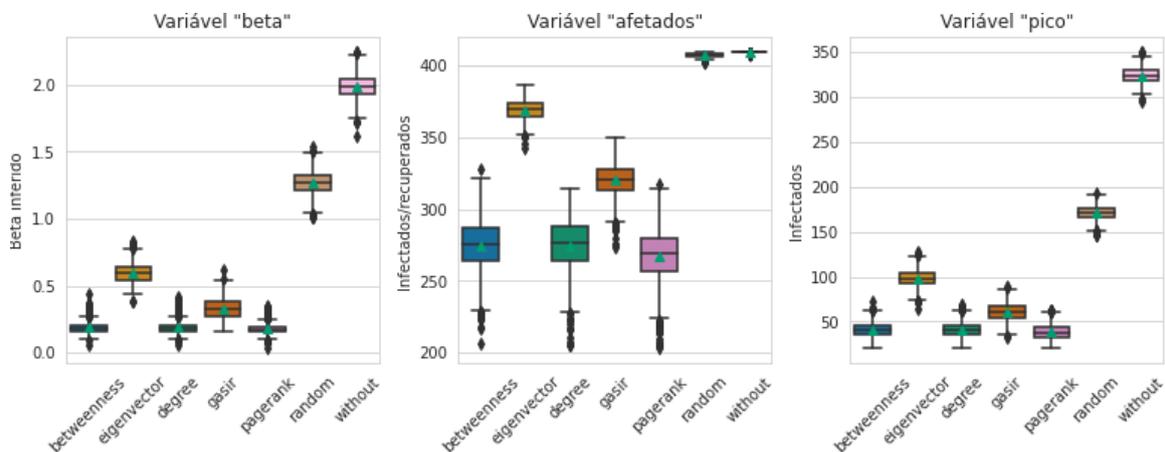
Figura 22 – Percentual da população que se manteve suscetível após a aplicação dos critérios de vacinação nas redes artificiais. Valor estimado a partir da mediana dos experimentos.

As Figuras 23 e 24 ilustram, em duas redes artificiais, valores nas variáveis dependentes muito próximos entre os critérios C_{pv}^B , C_{pv}^D e C_{pv}^P , embora haja equivalência estatística apenas entre os dois últimos. É possível visualizar pequenas diferenças de mediana e amplitude interquartil.



Fonte: O autor

Figura 23 – Boxplots com resultados das três variáveis dependentes com vacinação de 40% na rede *ba400-m3*.



Fonte: O autor

Figura 24 – Boxplots com resultados das três variáveis dependentes com vacinação de 40% na rede *ba400-m6*.

A Tabela 8 exibe os melhores critérios nas redes artificiais avaliados por sua mediana e classificados segundo cada uma das variáveis dependentes consideradas neste estudo.

Na análise das redes reais, vê-se comportamento semelhante dos critérios C_{pv}^P e C_{pv}^D nas redes artificiais. Eles apresentaram equivalência estatística diversas vezes nas redes *moreno highschool*, *moreno innovation* e *moreno oz*. O critério C_{pv}^B é semelhante somente na rede *moreno highschool*.

O número de indivíduos que se mantiveram suscetíveis nessas redes varia bastante e as maiores taxas observadas foram de 40% na rede *moreno highschool* e de 41.49% na rede *moreno innovation*. A Figura 25 ilustra que, nas redes *moreno oz* e *moreno health*, o percentual de indivíduos que se mantiveram suscetíveis após a simulação da epidemia são visualmente inferiores quando comparados às outras duas redes.

A Figura 26 mostra que na rede *moreno oz*, a variável de *afetados* apresentou uma contagem elevada de indivíduos, igual ou muito próxima do tamanho da rede. A rede, que

Tabela 8 – Melhores critérios avaliados pela mediana em redes artificiais

Rede	pv	Pico	Equivalentes	Beta	Equivalentes	Afetados	Equivalentes
ba400-m3	0.15	pagerank	\mathcal{C}^D	degree	\mathcal{C}^P	pagerank	
ba400-m3	0.20	degree	\mathcal{C}^P	degree	\mathcal{C}^P	degree	\mathcal{C}^P
ba400-m3	0.25	pagerank	\mathcal{C}^D	degree		pagerank	
ba400-m3	0.30	pagerank	\mathcal{C}^D	pagerank		pagerank	
ba400-m3	0.35	degree	\mathcal{C}^P	pagerank		pagerank	
ba400-m3	0.40	degree		degree		pagerank	
ba400-m4	0.15	degree	\mathcal{C}^P	degree	\mathcal{C}^P	pagerank	\mathcal{C}^D
ba400-m4	0.20	pagerank	\mathcal{C}^D	pagerank	\mathcal{C}^D	degree	\mathcal{C}^P
ba400-m4	0.25	degree	\mathcal{C}^P	degree	\mathcal{C}^P	pagerank	
ba400-m4	0.30	pagerank	\mathcal{C}^D	degree		pagerank	
ba400-m4	0.35	pagerank		pagerank		pagerank	
ba400-m4	0.40	degree	\mathcal{C}^P	degree	\mathcal{C}^P	pagerank	
ba400-m5	0.15	pagerank	$\mathcal{C}^D, \mathcal{C}^G$	degree	$\mathcal{C}^G, \mathcal{C}^P$	gasir	
ba400-m5	0.20	pagerank		pagerank		pagerank	\mathcal{C}^G
ba400-m5	0.25	pagerank		pagerank		degree	\mathcal{C}^P
ba400-m5	0.30	pagerank	\mathcal{C}^D	degree	\mathcal{C}^P	pagerank	\mathcal{C}^D
ba400-m5	0.35	pagerank		pagerank		pagerank	
ba400-m5	0.40	pagerank		pagerank		pagerank	
ba400-m6	0.15	gasir	$\mathcal{C}^P, \mathcal{C}^B, \mathcal{C}^D$	degree	$\mathcal{C}^G, \mathcal{C}^P, \mathcal{C}^B$	gasir	
ba400-m6	0.20	pagerank	\mathcal{C}^D	degree	\mathcal{C}^P	degree	
ba400-m6	0.25	pagerank	\mathcal{C}^B	betweenness	$\mathcal{C}^D, \mathcal{C}^P$	pagerank	\mathcal{C}^G
ba400-m6	0.30	pagerank		pagerank		pagerank	
ba400-m6	0.35	degree	\mathcal{C}^P	degree	\mathcal{C}^P	degree	\mathcal{C}^P
ba400-m6	0.40	pagerank		pagerank		pagerank	

possui o mais alto grau médio e um dos maiores coeficientes de agrupamento, cria a tendência de se atingir mais facilmente todos os indivíduos, o que dificulta em muito a contenção da epidemia. Contudo, as medianas e intervalos interquartis das variáveis *beta* e *pico* vistos nos *boxplots* demonstram que houve um decréscimo da intensidade da epidemia.

Na rede *moreno innovation* foi possível verificar que na variável *afetados* o critério \mathcal{C}_{pv}^B apresenta *outliers* com valores inferiores às medianas observadas nos critérios \mathcal{C}_{pv}^G , \mathcal{C}_{pv}^D e \mathcal{C}_{pv}^P , como ilustra a Figura 27. Isso representa que o ponto inicial de contágio na rede pode determinar que o critério \mathcal{C}_{pv}^B seja mais ou menos efetivo.

Os diferentes níveis de complexidade topológica das redes demonstra que as estruturas impactam no sucesso dos critérios de vacinação, fator que determina a importância da análise de outras características do espalhamento, como a taxa de contágio e o pico de infectados.

Na rede *moreno health* observou-se que os experimentos não apresentaram indícios de equivalência estatística entre os critérios aplicados. A maior fração da população mantida como suscetível foi de 15% e a maioria dos melhores resultados foram entregues pelo critério \mathcal{C}_{pv}^P , como pode ser visto na Figura 25. Como essa é a maior das redes, vale ressaltar que os critérios \mathcal{C}_{pv}^B , \mathcal{C}_{pv}^D e \mathcal{C}_{pv}^G , embora não sejam equivalentes, também são

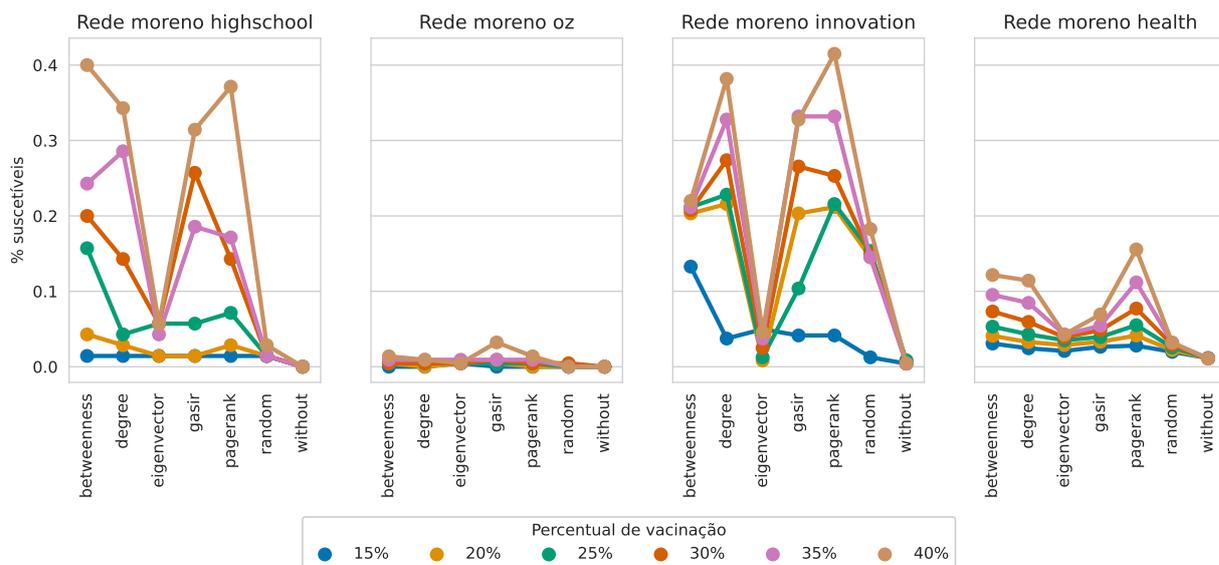


Figura 25 – Percentual da população que se manteve suscetível após a aplicação dos critérios de vacinação nas redes reais. Valor estimado a partir da mediana dos experimentos.

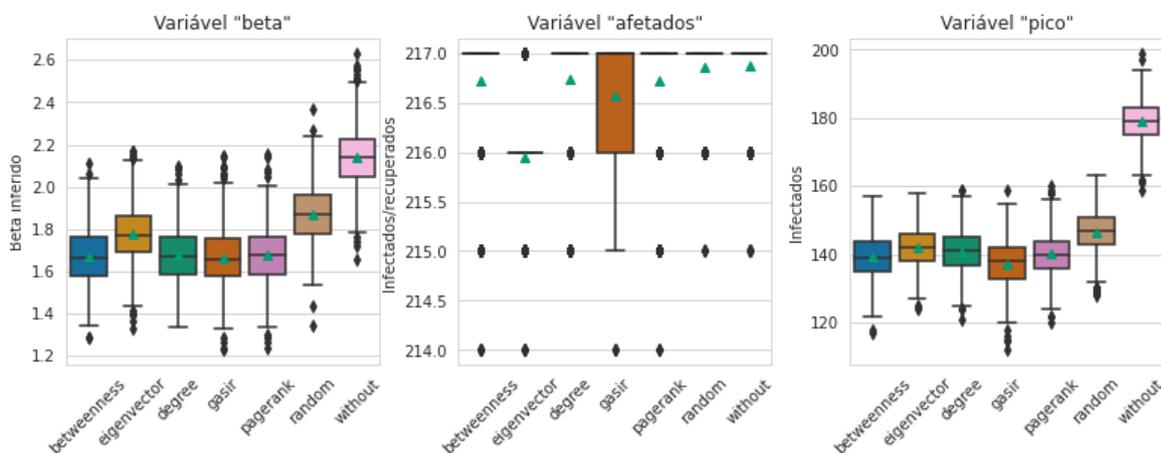
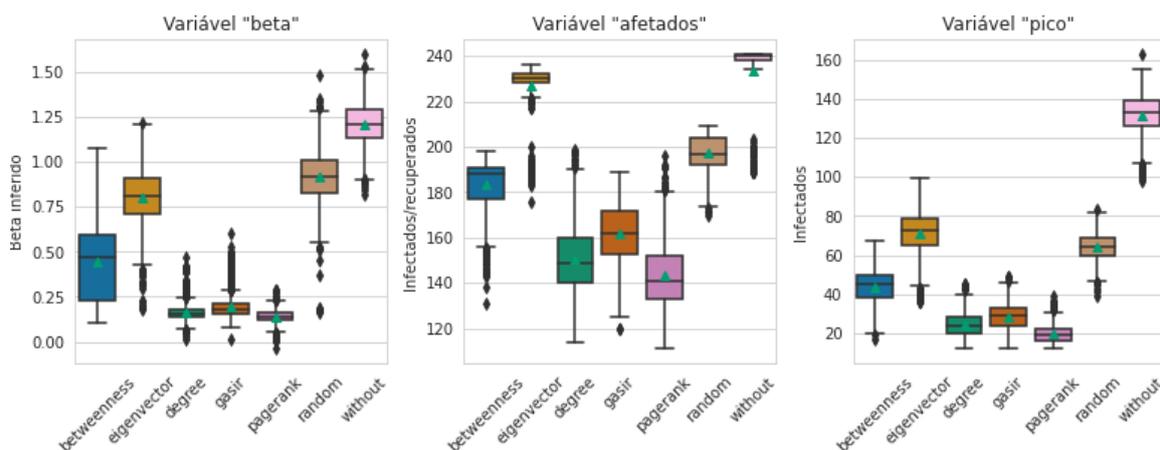


Figura 26 – Boxplots com resultados das três variáveis dependentes com vacinação de 15% na rede oz.

competentes na contenção e desaceleração da epidemia.

A Tabela 9 exhibe os melhores critérios nas redes reais avaliados por sua mediana e classificados segundo cada uma das variáveis dependentes consideradas neste estudo.

É possível concluir que, em todas as redes estudadas, os resultados obtidos pelo critério C_{pv}^G são bons e estão normalmente a margem dos critérios C_{pv}^P e C_{pv}^D . As Tabelas 11, 12, e 13 indicam sua seleção como o melhor critério em algum momento, mas em muitos desses cenários existe equivalência estatística com outro critério baseado em centralidade. Outra observação importante é que o critério C_{pv}^E é, dos métodos baseados em centralidade, o mais específico. Normalmente seus resultados foram piores que os demais e em muitas vezes seu desempenho se assemelha ao critério C_{pv}^R .



Fonte: O autor

Figura 27 – Boxplots com resultados das três variáveis dependentes com vacinação de 40% na rede *innovation*.

Tabela 9 – Melhores critérios avaliados pela mediana em redes reais.

Rede	pv	Pico	Equivalentes	Beta	Equivalentes	Afetados	Equivalentes
health	0.15	pagerank		pagerank		betweenness	
health	0.20	pagerank		pagerank		pagerank	C^B
health	0.25	pagerank		pagerank		pagerank	
health	0.30	pagerank		pagerank		pagerank	
health	0.35	pagerank		pagerank		pagerank	
health	0.40	pagerank		pagerank		pagerank	
highschool	0.15	degree	C^E, C^G, C^P	gasir	C^P, C^D	betweenness	C^D, C^G, C^P
highschool	0.20	pagerank		pagerank		betweenness	
highschool	0.25	betweenness	C^D, C^G, C^P	gasir	C^P, C^D	betweenness	
highschool	0.30	gasir	C^B	betweenness	C^G	gasir	C^B
highschool	0.35	degree	C^B	degree	C^B	degree	
highschool	0.40	betweenness	C^P	betweenness	C^P	betweenness	C^P
innovation	0.15	gasir	C^P, C^D	degree	C^G, C^P	betweenness	C^D, C^G, C^E, C^P
innovation	0.20	degree	C^P	degree	C^P	degree	C^P
innovation	0.25	gasir		gasir		degree	C^B
innovation	0.30	degree	C^G, C^P	degree	C^G, C^P	degree	C^G, C^P
innovation	0.35	pagerank		gasir	C^P, C^D	gasir	C^P, C^D
innovation	0.40	pagerank		pagerank		pagerank	
oz	0.15	gasir		betweenness	C^D, C^G, C^P	eigenvector	
oz	0.20	gasir	C^B	gasir	C^P, C^B, C^D	betweenness	
oz	0.25	gasir		gasir	C^P	betweenness	C^E
oz	0.30	gasir		pagerank	C^G	gasir	
oz	0.35	betweenness	C^G, C^P	gasir	C^P, C^B	betweenness	C^E
oz	0.40	gasir		pagerank	C^B	gasir	

Por fim, a utilização de uma escolha sistemática de indivíduos é comprovadamente eficaz, dada as diferenças de amplitude interquartil e medianas observadas nos critérios baseados em centralidade e no GASIR, comparativamente com os critérios aleatório e com a ausência de vacinação. A Figuras do Apêndice C ilustram detalhadamente essas diferenças.

4.3 Considerações finais do capítulo

Este capítulo elaborou e analisou um conjunto de experimentos que buscava averiguar a eficácia de diferentes critérios de vacinação em diferentes redes complexas.

A primeira investigação considerou analisar o quão equivalentes são os conjuntos de indivíduos selecionados pelos critérios em cada percentual de vacinação avaliado. Utilizou-se a medida de similaridade *rbo* e o número de indivíduos comuns entre os conjuntos. Os indivíduos selecionados pela centralidade *pagerank* e grau apresentaram constantemente um alto número de indivíduos comuns, assim como níveis de similaridade superiores a 0.5. Diversos indivíduos selecionados pela metodologia GASIR são comuns aos escolhidos por esses critérios de centralidade, mas o nível de similaridade é mais próximo de zero. Isso indica que o GASIR avaliou a importância dos seus indivíduos de forma distinta dos critérios *pagerank* e grau. Foi possível concluir que, normalmente, o GASIR apresenta resultados equivalentes a essas duas centralidades em simulações de contágio na rede.

Especificamente nas redes artificiais, viu-se que as medidas baseadas em centralidade apresentaram níveis de similaridades bastante altos.

Viu-se que os critérios *betweenness* e *eigenvector* possuem intersecção e níveis de similaridades maiores apenas em algumas redes, evidência de que estas centralidades podem apresentar resultados mais situacionais, conforme a estrutura topológica da rede.

Algumas das observações dos experimentos de contágio nas redes foram baseadas na análise estatística não paramétrica das amostras. Tal decisão decorreu do fato de que as amostras agrupadas por rede, percentual e critério de vacinação não seguiam uma distribuição normal. Como as técnicas de remoção de *outliers* ou de transformação das amostras não foram efetivas, optou-se por utilizar essa abordagem.

A estrutura topológica da rede possui um papel importante no processo de disseminação, visto que as variáveis dependentes estudadas podem apresentar melhores resultados com critérios diferentes para a mesma rede e percentual de vacinação. Em redes de maior complexidade viu-se que minimizar o número de indivíduos afetados pode ser um trabalho difícil, mas que é possível reduzir a taxa de contágio, como visto nos resultados providos pelas variáveis dependentes *beta* e *pico*.

Por fim, observou-se que o critério que utiliza a centralidade *pagerank* apresentou com maior frequência os melhores resultados na avaliação das variáveis dependentes na maioria das situações experimentadas. Contudo, os resultados equivalentes entre os critérios *pagerank* e grau demonstram que o último, que demanda baixo custo computacional para cálculo, pode ser uma alternativa viável para redes muito grandes.

A análise comprovou que, sob o ponto de vista das três variáveis dependentes analisadas, os resultados obtidos pela metodologia GASIR são, muitas vezes, similares aos

critérios que utilizavam a centralidade por grau e *pagerank*. Tal comportamento corroborou com a expectativa gerada pela análise da similaridade e da intersecção dos indivíduos selecionados pelos critérios. Com isso não se viu vantagem competitiva para a metodologia GASIR, devido ao seu alto custo computacional para a escolha dos indivíduos. É possível que a metodologia apresente desempenho diferenciado em redes com características topológicas mais complexas, o que seria desejável em trabalhos futuros.

5 Conclusões

Em cenários de recursos limitados para imunização passiva, a escolha assertiva de indivíduos a serem vacinados pode minimizar o número de pessoas afetadas pela doença ou ainda diminuir a velocidade de contágio. Um número restrito de doses pode caracterizar um problema combinatório bastante difícil de se solucionar. A utilização do conhecimento de redes complexas e de modelos epidemiológicos cria um cenário favorável para a construção de hipóteses que podem facilitar as tomadas de decisão.

Com auxílio desse ferramental este trabalho propôs a metodologia de otimização GASIR, baseada na estratégia de mesmo nome apresentada no trabalho de [Rodrigues \(2018\)](#). Seu objetivo é a escolha de indivíduos ideais a serem vacinados em uma rede complexa a partir dos parâmetros da doença e com auxílio de um algoritmo de otimização. O grau de eficácia da metodologia foi avaliada comparativamente com outros critérios de escolha de indivíduos que utilizaram de medidas de centralidade para a seleção.

A análise comparativa da ausência e de diferentes critérios de vacinação indicou que uma escolha sistemática tende a reduzir o número de infectados ou a taxa de contágio em uma rede complexa. O critério de seleção baseado na metodologia GASIR apresentou resultados próximos de outros baseados em centralidade, mas a um custo computacional bastante superior.

Na metodologia de otimização proposta aplicou-se ajustes na função objetivo para a melhoria da sua efetividade na avaliação de grupos de vacinação, realizou-se ajustes em diversas operações genéticas e investigou-se os parâmetros do AG. Foi verificado um nível satisfatório de estabilidade da metodologia, observada pela melhor acurácia em testes de contágio que consideram a origem da infecção e pela habilidade do modelo em propor conjuntos-solução semelhantes entre diferentes execuções.

Por fim, o conjunto de experimentos avaliativos propostos mostrou que os níveis de similaridade e intersecção dos critérios que utilizam a centralidade de grau e centralidade de *pagerank* eram frequentemente equivalentes. Esses critérios apresentaram os melhores resultados com frequência maior do que os demais. Outras medidas de centralidade, como o *betweenness* e o *eigenvector* apresentaram alguma similaridade às anteriores em redes com topologias mais específicas, especialmente em redes que possuíam maior coeficiente de agrupamento.

Viu-se que, no cenário estudado, a metodologia GASIR desempenhou-se bem no seu propósito, mas não se verificou vantagem competitiva ante os critérios de centralidade que apresentaram os melhores resultados. Viu-se que seus conjuntos selecionados são frequentemente equivalentes aos selecionados pelos critérios de centralidade por grau e

pagerank e sem grandes diferenças nos experimentos simulados de epidemia.

5.1 Limitações e trabalhos futuros

A metodologia GASIR apresenta alto custo computacional, especialmente pela natureza estocástica da sua função objetivo. Sua convergência prematura ou a tendência de se contaminar toda a rede geram a necessidade de técnicas de reparação da avaliação de soluções para que o algoritmo de otimização consiga desempenhar o seu papel. Tudo isso inviabiliza a sua aplicação em redes maiores.

Uma das alternativas vislumbradas é que a utilização de algoritmos não populacionais, que considerem aspectos como a centralidade dos vértices para aplicação em heurísticas construtivas, possa trazer maior agilidade e estabilidade na proposição dos indivíduos ideais. Outra possibilidade é a substituição da função objetivo por um meta modelo, capaz de avaliar a escolha dos indivíduos segundo suas características topológicas na rede, como coeficiente de agrupamento, número de vértices ou sua avaliação segundo alguma centralidade. Ambas alternativas considerariam as características topológicas da rede estudada, fator importante na disseminação de doenças em uma rede conforme observado na análise dos critérios avaliados neste trabalho.

A utilização do modelo epidemiológico SIR estocástico como ferramenta comparativa de diferentes critérios de vacinação limitou a discussão dos resultados sob perspectiva não-paramétrica dos grupos de amostras. Trabalhos futuros podem expandir a discussão dessas abordagens, no anseio de se consolidar este método avaliativo.

Seria importante ainda uma maior diversidade de ambientes experimentados. A utilização de modelos de redes complexas que reproduzam ambientes urbanos, como abordado em [Rui \(2013\)](#) e [Rodrigues et al. \(2018\)](#), poderiam elaborar cenários de maior complexidade topológica para avaliação. A utilização desses cenários junto ao SIR aplicado em redes complexas pode desencadear percepções mais detalhadas entre a organização espacial da população e a disseminação da doença, características que são generalizadas em estudos que utilizam modelo SIR clássico para a análise de epidemias.

Referências

- ROUQUAYROL, M. Z.; FILHO, N. d. A. *Epidemiologia & saúde*. [S.l.]: MEDSI, 2003. ISBN 978-85-7199-351-8. Citado na página 21.
- PORTA, M. *A Dictionary of Epidemiology*. [S.l.]: Oxford University Press, 2014. ISBN 978-0-19-939005-2. Citado na página 21.
- FEIJO, R. B.; SAFADI, M. A. P. Imunizações: três séculos de uma história de sucessos e constantes desafios. *Jornal de Pediatria*, Sociedade Brasileira de Pediatria, v. 82, n. 3, p. s1–s3, jul. 2006. ISSN 0021-7557. Disponível em: <<http://www.scielo.br>>. Citado na página 21.
- CDC, N. C. f. H. S. *A CDC framework for preventing infectious diseases; sustaining the essentials and innovating for the future*. Atlanta, GA: [s.n.], 2011. Report. Disponível em: <<https://stacks.cdc.gov/view/cdc/11695>>. Citado na página 21.
- LI, X.; MUKANDAVIRE, C.; CUCUNUBÁ, Z. M.; LONDONO, S. E.; ABBAS, K.; CLAPHAM, H. E.; JIT, M.; JOHNSON, H. L.; PAPADOPOULOS, T.; VYNNYCKY, E. Estimating the health impact of vaccination against ten pathogens in 98 low-income and middle-income countries from 2000 to 2030: A modelling study. *The Lancet*, Elsevier, v. 397, n. 10272, p. 398–408, 2021. Citado na página 21.
- RODRIGUES, H. S. Application of sir epidemiological model: New trends. *arXiv:1611.02565 [physics, q-bio]*, nov. 2016. Disponível em: <<http://arxiv.org/abs/1611.02565>>. Citado na página 21.
- KEMPE, D.; KLEINBERG, J.; TARDOS, E. Influential nodes in a diffusion model for social networks. In: CAIRES, L.; ITALIANO, G. F.; MONTEIRO, L.; PALAMIDESSI, C.; YUNG, M. (Ed.). *Automata, Languages and Programming*. Berlin, Heidelberg: Springer, 2005. (Lecture Notes in Computer Science), p. 1127–1138. ISBN 978-3-540-31691-6. Citado na página 22.
- OLIVEIRA, R. S.; XAVIER, C. R.; VIEIRA, V. d. F.; REIS, R. F.; SANTOS, R. W. dos; QUINTELA, B. d. M.; ROCHA, B. M. *Nota Técnica 04: Como Um Esquema Sólido de Vacinação Pode Controlar a Pandemia de COVID-19 No Brasil?* [S.l.], fev. 2021. Disponível em: <https://www.ufjf.br/pgmc/files/2020/03/nt04_covid19_pgmc_ufjf_ufsj.pdf>. Citado na página 23.
- NDIAYE, B. M.; TENDENG, L.; SECK, D. Comparative prediction of confirmed cases with covid-19 pandemic by machine learning, deterministic and stochastic sir models. *arXiv:2004.13489 [math, q-bio, stat]*, abr. 2020. Disponível em: <<http://arxiv.org/abs/2004.13489>>. Citado 2 vezes nas páginas 24 e 69.
- REIS, R. F.; de Melo Quintela, B.; de Oliveira Campos, J.; GOMES, J. M.; POMPEI, L. T.; ROCHA, B. M.; LOBOSCO, M.; SANTOS, R. W. dos. *Características Da Pandemia Do Covid-19 e Impacto Das Políticas de Isolamento No Brasil, Minas Gerais e Juiz de Fora*. [S.l.], mar. 2020. Disponível em: <<http://www.ufjf.br/pgmc/files/2020/03/nt01pgmc-30-03-2020.pdf>>. Citado na página 24.

- REIS, R. F.; OLIVEIRA, R. S.; XAVIER, C. R.; SANTOS, R. W. dos. *Nota Técnica 03: Impacto Das Mudanças de Políticas de Mitigação Ao Longo Do Tempo: Caracterização e Projeções Para a Pandemia COVID-19 Do Brasil, Minas Gerais, Juiz de Fora e São João Del-Rei*. [S.l.], jul. 2020. Disponível em: <https://www.ufjf.br/pgmc/files/2020/03/nt03_covid19_pgmc_ufjf2.pdf>. Citado na página 24.
- TANNER, M. W.; SATTENSPIEL, L.; NTAIMO, L. Finding optimal vaccination strategies under parameter uncertainty using stochastic programming. *Mathematical Biosciences*, v. 215, n. 2, p. 144–151, out. 2008. ISSN 00255564. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0025556408001156>>. Citado na página 25.
- SHULGIN, B.; STONE, L.; AGUR, Z. Pulse vaccination strategy in the sir epidemic model. *Bulletin of Mathematical Biology*, v. 60, n. 6, p. 1123–1148, nov. 1998. ISSN 1522-9602. Disponível em: <[https://doi.org/10.1006/S0092-8240\(98\)90005-2](https://doi.org/10.1006/S0092-8240(98)90005-2)>. Citado na página 25.
- NGUYEN, C.; CARLSON, J. M. Optimizing real-time vaccine allocation in a stochastic sir model. *PLoS ONE*, v. 11, n. 4, abr. 2016. ISSN 1932-6203. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4820222/>>. Citado na página 25.
- LIU, L.; LUO, X.; CHANG, L. Vaccination strategies of an sir pair approximation model with demographics on complex networks. *Chaos, Solitons & Fractals*, v. 104, p. 282–290, nov. 2017. ISSN 0960-0779. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S096007791730348X>>. Citado na página 25.
- LI, Z.; ZHU, P.; ZHAO, D.; DENG, Z.; WANG, Z. Suppression of epidemic spreading process on multiplex networks via active immunization. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, v. 29, n. 7, p. 073111, jul. 2019. ISSN 1054-1500. Disponível em: <<https://aip.scitation.org/doi/10.1063/1.5093047>>. Citado na página 25.
- ABBASSI, Z.; HEIDARI, H. Toward optimal vaccination strategies for probabilistic models. In: *Proceedings of the 20th International Conference Companion on World Wide Web - WWW '11*. Hyderabad, India: ACM Press, 2011. p. 1. ISBN 978-1-4503-0637-9. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1963192.1963194>>. Citado na página 25.
- PASTOR-SATORRAS, R.; CASTELLANO, C.; MIEGHEM, P. V.; VESPIGNANI, A. Epidemic processes in complex networks. *Reviews of Modern Physics*, v. 87, n. 3, p. 925–979, ago. 2015. ISSN 0034-6861, 1539-0756. Disponível em: <<https://link.aps.org/doi/10.1103/RevModPhys.87.925>>. Citado 5 vezes nas páginas 25, 32, 35, 37 e 38.
- WOO, J.; SON, J.; CHEN, H. An sir model for violent topic diffusion in social media. In: *Proceedings of 2011 IEEE International Conference on Intelligence and Security Informatics*. [S.l.: s.n.], 2011. p. 15–19. ISSN null. Citado na página 25.
- KANDHWAY, K.; KURI, J. How to run a campaign: Optimal control of sis and sir information epidemics. *Applied Mathematics and Computation*, v. 231, p. 79–92, mar. 2014. ISSN 00963003. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0096300314000022>>. Citado na página 25.

- ERDOS, P.; RÉNYI, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, v. 5, n. 1, p. 17–60, 1960. Citado na página 27.
- GRANOVETTER, M. S. The strength of weak ties. In: *Social Networks*. [S.l.]: Elsevier, 1977. p. 347–367. Citado na página 27.
- BARABASI, A.-L. *Network Science*. [s.n.], 2016. ISBN 978-1-107-07626-6. Disponível em: <<http://networksciencebook.com/>>. Citado 4 vezes nas páginas 27, 29, 33 e 36.
- NEWMAN, M. E. The structure and function of complex networks. *SIAM review*, v. 45, n. 2, p. 167–256, 2003. Citado 3 vezes nas páginas 27, 30 e 32.
- BRITO, S. G. d. A. *Contribuição Ao Estudo de Redes Complexas: Modelo de Afinidade Com Métrica*. Tese (Master's Thesis) — Universidade Federal do Rio Grande do Norte, 2012. Citado na página 27.
- FAGIOLO, G. Clustering in complex directed networks. *Physical Review E*, APS, v. 76, n. 2, p. 026107, 2007. Citado na página 29.
- NEWMAN, M. E. Assortative mixing in networks. *Physical review letters*, APS, v. 89, n. 20, p. 208701, 2002. Citado 2 vezes nas páginas 31 e 65.
- P. Erdos; RÉNYI, A. On random graphs i. *Publ. math. debrecen*, v. 6, n. 290-297, p. 18, 1959. Citado na página 31.
- GILBERT, E. N. Random graphs. *The Annals of Mathematical Statistics*, JSTOR, v. 30, n. 4, p. 1141–1144, 1959. Citado na página 31.
- WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. *nature*, Nature Publishing Group, v. 393, n. 6684, p. 440–442, 1998. Citado na página 31.
- BARABASI, A.-L.; ALBERT, R. Emergence of scaling in random networks. *science*, American Association for the Advancement of Science, v. 286, n. 5439, p. 509–512, 1999. Citado 2 vezes nas páginas 32 e 64.
- PRICE, D. D. S. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, v. 27, n. 5, p. 292–306, set. 1976. ISSN 00028231, 10974571. Disponível em: <<http://doi.wiley.com/10.1002/asi.4630270505>>. Citado na página 32.
- FREEMAN, L. C. Centrality in social networks conceptual clarification. v. 1, n. 3, p. 215–239, 1978. Citado na página 33.
- BONACICH, P. Power and centrality: A family of measures the american journal of sociology, vol. 92. 1987. Citado na página 33.
- PAGE, L.; BRIN, S.; MOTWANI, R.; WINOGRAD, T. *The PageRank Citation Ranking: Bringing Order to the Web*. [S.l.], 1999. Citado na página 34.
- BERNOULLI, D. Essai d’une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l’inoculation pour la prévenir. *Histoire de l’Acad., Roy. Sci.(Paris) avec Mem*, p. 1–45, 1760. Citado na página 34.

- ANDERSON, R. M.; MAY, R. M. Population biology of infectious diseases: Part i. *Nature*, Nature Publishing Group, v. 280, n. 5721, p. 361–367, ago. 1979. ISSN 1476-4687. Disponível em: <<https://www.nature.com/articles/280361a0>>. Citado na página 34.
- ANDERSON, R. M.; MAY, R. M. *Infectious Diseases of Humans: Dynamics and Control*. Oxford ; New York: Oxford University Press, 1991. (Oxford Science Publications). ISBN 978-0-19-854599-6. Citado na página 34.
- KEELING, M. J.; ROHANI, P. *Modeling Infectious Diseases in Humans and Animals*. Princeton: Princeton University Press, 2008. ISBN 978-0-691-11617-4. Citado 2 vezes nas páginas 34 e 36.
- BRAUER, F.; Castillo-Chávez, C. *Mathematical Models in Population Biology and Epidemiology*. 2nd ed. ed. New York: Springer, 2012. (Texts in Applied Mathematics, 40). ISBN 978-1-4614-1685-2. Citado na página 34.
- BAILEY, N. T. J. *The Mathematical Theory of Infectious Diseases and Its Applications*. 2nd ed. ed. London: Griffin, 1975. ISBN 978-0-85264-231-3. Citado na página 34.
- KERMACK, W. O.; MCKENDRICK, A. G.; WALKER, G. T. 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*, v. 115, n. 772, p. 700–721, ago. 1927. Disponível em: <<https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1927.0118>>. Citado 2 vezes nas páginas 35 e 36.
- DIETZ, K. Epidemics and rumours: A survey. *Journal of the Royal Statistical Society. Series A (General)*, v. 130, n. 4, p. 505, 1967. ISSN 00359238. Disponível em: <<https://www.jstor.org/stable/10.2307/2982521?origin=crossref>>. Citado na página 35.
- PEIXOTO, T. P. *graph-tool.dynamics - Dynamical processes documentation*. [S.l.], 2020. Disponível em: <<https://graph-tool.skewed.de/static/doc/dynamics.html>>. Citado na página 38.
- BLUM, C.; ROLI, A. Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys*, v. 35, n. 3, p. 268–308, set. 2003. ISSN 03600300. Disponível em: <<http://portal.acm.org/citation.cfm?doid=937503.937505>>. Citado na página 39.
- GOLDBERG, D. E.; ZAMES, G.; AJLOUNI, N. M.; AJLOUNI, N. M.; AJLOUNI, N. M.; HOLLAND, J. H.; HILLS, W. D. Genetic algorithms in search, optimization and machine learning. *Information Technology Journal*, v. 3, n. 1, p. 301–302, 1981. Citado 2 vezes nas páginas 40 e 43.
- HOLLAND, J. H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Ann Arbor: University of Michigan Press, 1975. ISBN 978-0-472-08460-9. Citado 2 vezes nas páginas 40 e 43.
- BACK, T. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. New York: Oxford University Press, 1996. ISBN 978-0-19-509971-3. Citado na página 40.

- TALBI, E.-G. *Metaheuristics: From Design to Implementation*. Hoboken, N.J.: John Wiley & Sons, 2009. ISBN 978-0-470-27858-1. Citado na página 42.
- MÜHLENBEIN, H.; SCHLIERKAMP-VOOSEN, D. Predictive models for the breeder genetic algorithm i. continuous parameter optimization. *Evolutionary Computation*, MIT Press, v. 1, n. 1, p. 25–49, mar. 1993. ISSN 1063-6560. Disponível em: <<https://doi.org/10.1162/evco.1993.1.1.25>>. Citado na página 43.
- BLICKLE, T.; THIELE, L. *A Comparison of Selection Schemes Used in Genetic Algorithms*. 1995. Disponível em: </paper/A-Comparison-of-Selection-Schemes-used-in-Genetic-Blickle-Thiele/a5532dda955228ea44e2d224c6b42916959705b1>. Citado na página 43.
- LIM, S. M.; SULTAN, A. B. M.; SULAIMAN, M. N.; MUSTAPHA, A.; LEONG, K. Y. Crossover and mutation operators of genetic algorithms. *International journal of machine learning and computing*, International Journal of Machine Learning and Computing (IJMLC), v. 7, n. 1, p. 9–12, 2017. Citado na página 44.
- DRÉO, J.; PÉTROWSKI, A.; SIARRY, P.; TAILLARD, E. *Metaheuristics for Hard Optimization: Methods and Case Studies*. 2006 edition. ed. [S.l.]: Springer, 2006. Citado na página 45.
- SYSWERDA, G. A study of reproduction in generational and steady-state genetic algorithms. In: *Foundations of Genetic Algorithms*. Elsevier, 1991. v. 1, p. 94–101. ISBN 978-0-08-050684-5. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/B9780080506845500094>>. Citado na página 45.
- LOUREIRO, M. M. S. *Um método de reanálise adaptativa para otimização estrutural usando um algoritmo genético não-geracional*. Tese (@phdthesis) — Universidade Federal de Juiz de Fora, Juiz de Fora, fev. 2016. Disponível em: <<https://repositorio.ufjf.br/jspui/handle/ufjf/1393>>. Citado na página 45.
- BABU, T. R. *Compression Schemes for Mining Large Datasets: A Machine Learning Perspective*. 1st edition. ed. New York: Springer, 2013. (Advances in Computer Vision and Pattern Recognition). ISBN 978-1-4471-5606-2. Citado na página 45.
- JONG, K. A. D. *Analysis of the Behavior of a Class of Genetic Adaptive Systems*. Tese (Doutorado) — University of Michigan, 1975. Citado na página 45.
- JONG, K. A. D.; SARMA, J. Generation gaps revisited. In: WHITLEY, L. D. (Ed.). *Foundations of Genetic Algorithms*. Elsevier, 1993, (Foundations of Genetic Algorithms, v. 2). p. 19–28. Disponível em: <<http://www.sciencedirect.com/science/article/pii/B9780080948324500076>>. Citado na página 45.
- RODRIGUES, R. F. *GASIR*. 2018. Disponível em: <<https://github.com/RodrigoFerreira001/GASIR>>. Citado 2 vezes nas páginas 46 e 50.
- PEIXOTO, T. P. The graph-tool python library. *figshare*, 2014. Disponível em: <http://figshare.com/articles/graph_tool/1164194>. Citado na página 47.
- BRESEE, J. S.; FIORE, A. E.; FRY, A.; GUBAREVA, L. V.; SHAY, D. K.; UYEKI, T. M. Antiviral agents for the treatment and chemoprophylaxis of influenza: Recommendations of the advisory committee on immunization practices (acip). *Morbidity*

- and Mortality Weekly Report*, v. 60, p. 1–24, jan. 2011. Citado 2 vezes nas páginas 49 e 66.
- COBURN, B. J.; WAGNER, B. G.; BLOWER, S. Modeling influenza epidemics and pandemics: Insights into the future of swine flu (h1n1). *BMC Medicine*, v. 7, p. 30, jun. 2009. ISSN 1741-7015. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2715422/>>. Citado 2 vezes nas páginas 49 e 66.
- RODRIGUES, R. F. *Escolha de Indivíduos a Serem Imunizados via Algoritmo Genético Sobre o Modelo SIR Em Redes Complexas*. Monografia (Bacharelado) — Universidade Federal de São João Del Rei, nov. 2018. Citado 2 vezes nas páginas 50 e 81.
- PEIXOTO, T. P. *The Netzschleuder Network Catalogue and Repository*. 2020. Disponível em: <<https://networks.skewed.de/>>. Citado na página 64.
- COLEMAN, J. S. *Introduction to Mathematical Sociology*. [S.l.]: London Free Press Glencoe, 1964. Citado na página 64.
- COLEMAN, J.; KATZ, E.; MENZEL, H. The diffusion of an innovation among physicians. *Sociometry*, v. 20, n. 4, p. 253, dez. 1957. ISSN 00380431. Disponível em: <<https://www.jstor.org/stable/2785979?origin=crossref>>. Citado na página 64.
- FREEMAN, L. C.; WEBSTER, C. M.; KIRKE, D. M. Exploring social structure using dynamic three-dimensional color images. *Social Networks*, v. 20, n. 2, p. 109–118, abr. 1998. ISSN 03788733. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0378873397000166>>. Citado na página 65.
- MOODY, J. Peer influence groups: Identifying dense clusters in large networks. *Social Networks*, v. 23, n. 4, p. 261–283, out. 2001. ISSN 03788733. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0378873301000429>>. Citado na página 65.
- HELD, P.; DOCKHORN, A.; KRUSE, R. On merging and dividing of barabasi-albert-graphs. In: *2014 IEEE Symposium on Evolving and Autonomous Learning Systems (EALS)*. [S.l.]: IEEE, 2014. p. 17–24. Citado na página 65.
- WEBBER, W.; MOFFAT, A.; ZOBEL, J. A similarity measure for indefinite rankings. *ACM Transactions on Information Systems (TOIS)*, ACM New York, NY, USA, v. 28, n. 4, p. 1–38, 2010. Citado na página 67.
- HARTVIGSEN, G.; DRESCH, J. M.; ZIELINSKI, A. L.; MACULA, A. J.; LEARY, C. C. Network structure, and vaccination strategy and effort interact to affect the dynamics of influenza epidemics. *Journal of Theoretical Biology*, v. 246, n. 2, p. 205–213, maio 2007. ISSN 0022-5193. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022519306005972>>. Citado na página 70.
- SHAPIRO, S.; WILK, M. An analysis of variance test for normality (complete samples). *Biometrika*, v. 52, n. 3/4, p. 591–611, 1965. Citado na página 70.
- BOX, G.; COX, D. An analysis of transformations”. *Journal of the Royal Statistical Society B*, v. 26, p. 211–252, 1964. Citado na página 71.

- OLIVEIRA, J. P. R.; SANTANA, D. G. de; PEREIRA, V. J.; SANTOS, C. M. dos; Ribeiro-Oliveira, J. P.; SANTANA, D. G. de; PEREIRA, V. J.; SANTOS, C. M. dos. Data transformation: An underestimated tool by inappropriate use. *Acta Scientiarum. Agronomy*, Editora da Universidade Estadual de Maringá - EDUEM, v. 40, 2018. ISSN 1807-8621. Disponível em: <<http://www.scielo.br>>. Citado na página 71.
- KRUSKAL, W.; WALLIS, W. Use of ranks in one-criterion variance analysis". *Journal of the American Statistical Association*, v. 47, n. ue 260, p. 583–621,, 1952. Citado na página 71.
- MANN, H.; WHITNEY, D. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, p. 50–60, 1947. Citado na página 71.
- RUI, Y. *Urban growth modeling based on land-use changes and road network expansion*. Tese (Doutorado) — KTH Royal Institute of Technology, 2013. Citado na página 82.
- RODRIGUES, F.; XAVIER, C. R.; EVSUKOFF, A. G.; ZIVIANI, A.; da F. Vieira, V. Análise de mobilidade urbana em diferentes estratos sociais a partir de uma rede de telefonia móvel. In: *Anais Do II Workshop de Computação Urbana*. Porto Alegre, RS, Brasil: SBC, 2018. ISSN 2595-2706. Disponível em: <<https://sol.sbc.org.br/index.php/courb/article/view/2333>>. Citado na página 82.

Apêndices

APÊNDICE A – Tabelas com resultados de alguns experimentos

Neste Capítulo estão agrupados todos os resultados maiores de experimentos. Estes foram colocados no apêndice para que a experiência de leitura não fosse comprometida.

Tabela 10 – Calibragem do AG após as mudanças propostas. Taxa de sucesso dada após 100 execuções de cada conjunto de condições aplicado ao *experimento base*.

p_{ag}	g_{ag}	m_{ag}	Sobrevivência	s_{ag}	Histograma 10%	Melhor solução
200	200	0.05	Posicional	-	1.00	0.32
200	200	0.05	Steady-State	1.0	1.00	0.00
200	200	0.05	Steady-State	0.9	1.00	0.01
200	200	0.05	Steady-State	0.8	1.00	0.00
200	200	0.05	Steady-State	0.7	1.00	0.01
200	200	0.05	Steady-State	0.6	1.00	0.02
200	200	0.05	Steady-State	0.5	1.00	0.05
200	100	0.05	Steady-State	1.0	0.99	0.02
200	100	0.05	Steady-State	0.9	0.99	0.05
200	200	0.10	Steady-State	0.8	0.99	0.00
200	100	0.05	Steady-State	0.8	0.99	0.04
200	100	0.05	Steady-State	0.7	0.99	0.01
200	200	0.10	Steady-State	0.5	0.99	0.02
100	200	0.05	Steady-State	1.0	0.99	0.00
100	200	0.05	Steady-State	0.9	0.99	0.00
200	200	0.05	Steady-State	0.4	0.98	0.05
200	100	0.05	Posicional	-	0.98	0.35
200	200	0.10	Steady-State	0.7	0.98	0.00
200	200	0.10	Steady-State	0.6	0.98	0.00
200	200	0.05	Steady-State	0.3	0.98	0.03
200	100	0.10	Steady-State	1.0	0.97	0.01
200	100	0.10	Steady-State	0.9	0.97	0.01
100	200	0.05	Steady-State	0.8	0.97	0.01
200	200	0.10	Steady-State	0.9	0.95	0.00
200	100	0.10	Steady-State	0.7	0.95	0.02
100	200	0.05	Steady-State	0.7	0.95	0.01
100	200	0.05	Steady-State	0.6	0.95	0.01
100	200	0.05	Steady-State	0.5	0.95	0.06
200	100	0.10	Steady-State	0.8	0.93	0.03
200	100	0.05	Steady-State	0.6	0.93	0.08
200	200	0.10	Steady-State	0.4	0.93	0.01

Continua ...

Tabela 10 – Calibragem do AG após as mudanças propostas. Taxa de sucesso dada após 100 execuções de cada conjunto de condições aplicado ao *experimento base*.

p_{ag}	g_{ag}	m_{ag}	Sobrevivência	s_{ag}	Histograma 10%	Melhor solução
200	200	0.10	Steady-State	1.0	0.92	0.00
100	100	0.05	Steady-State	0.9	0.92	0.07
200	50	0.05	Steady-State	1.0	0.91	0.09
200	100	0.05	Steady-State	0.5	0.91	0.10
200	100	0.10	Posicional	-	0.90	0.26
100	100	0.05	Steady-State	1.0	0.89	0.05
100	200	0.05	Steady-State	0.4	0.88	0.03
200	50	0.05	Steady-State	0.9	0.87	0.10
100	200	0.10	Steady-State	1.0	0.87	0.00
100	200	0.10	Steady-State	0.8	0.86	0.00
100	200	0.10	Steady-State	0.6	0.86	0.01
50	200	0.05	Steady-State	0.9	0.86	0.01
200	200	0.10	Steady-State	0.3	0.86	0.04
100	200	0.10	Steady-State	0.9	0.85	0.00
100	100	0.05	Steady-State	0.8	0.82	0.03
100	200	0.10	Steady-State	0.7	0.80	0.02
50	200	0.05	Steady-State	1.0	0.80	0.00
50	200	0.05	Steady-State	0.8	0.80	0.00
200	100	0.10	Steady-State	0.6	0.79	0.06
200	200	0.05	Steady-State	0.2	0.77	0.15
200	100	0.05	Steady-State	0.4	0.76	0.14
100	100	0.05	Steady-State	0.7	0.76	0.08
50	200	0.05	Steady-State	0.7	0.74	0.03
200	100	0.10	Steady-State	0.5	0.73	0.09
100	100	0.10	Steady-State	1.0	0.73	0.02
100	100	0.05	Posicional	-	0.72	0.39
200	50	0.05	Steady-State	0.8	0.71	0.19
100	100	0.05	Steady-State	0.6	0.71	0.04
100	200	0.10	Steady-State	0.5	0.71	0.06
100	100	0.10	Steady-State	0.9	0.65	0.02
200	50	0.10	Steady-State	1.0	0.62	0.06
200	50	0.05	Posicional	-	0.61	0.41
200	50	0.10	Steady-State	0.9	0.61	0.08
100	200	0.10	Steady-State	0.4	0.61	0.05
100	200	0.05	Steady-State	0.3	0.61	0.06
200	50	0.05	Steady-State	0.7	0.60	0.10
100	100	0.10	Steady-State	0.8	0.58	0.03
100	100	0.10	Steady-State	0.7	0.58	0.04
100	100	0.05	Steady-State	0.5	0.58	0.09
200	200	0.10	Steady-State	0.2	0.57	0.09
50	200	0.05	Steady-State	0.6	0.57	0.01

Continua ...

Tabela 10 – Calibragem do AG após as mudanças propostas. Taxa de sucesso dada após 100 execuções de cada conjunto de condições aplicado ao *experimento base*.

p_{ag}	g_{ag}	m_{ag}	Sobrevivência	s_{ag}	Histograma 10%	Melhor solução
50	100	0.05	Steady-State	1.0	0.55	0.03
50	200	0.05	Steady-State	0.5	0.55	0.03
200	100	0.10	Steady-State	0.4	0.52	0.09
100	50	0.05	Steady-State	1.0	0.52	0.10
100	200	0.05	Posicional	-	0.51	0.05
200	50	0.10	Steady-State	0.8	0.48	0.10
200	100	0.05	Steady-State	0.3	0.48	0.17
100	100	0.10	Steady-State	0.6	0.44	0.06
50	200	0.10	Steady-State	1.0	0.43	0.00
50	100	0.05	Steady-State	0.9	0.43	0.06
100	200	0.10	Steady-State	0.3	0.42	0.05
100	200	0.05	Steady-State	0.2	0.42	0.11
50	200	0.10	Steady-State	0.9	0.42	0.01
50	200	0.05	Steady-State	0.4	0.41	0.02
50	100	0.05	Steady-State	0.8	0.40	0.02
50	200	0.10	Steady-State	0.7	0.40	0.03
50	200	0.10	Steady-State	0.6	0.40	0.04
200	50	0.10	Steady-State	0.7	0.39	0.06
100	100	0.10	Posicional	-	0.38	0.12
50	200	0.10	Steady-State	0.5	0.37	0.02
100	50	0.05	Steady-State	0.9	0.36	0.05
100	100	0.05	Steady-State	0.4	0.36	0.08
100	50	0.10	Steady-State	1.0	0.35	0.08
50	100	0.05	Posicional	-	0.35	0.18
200	50	0.05	Steady-State	0.6	0.34	0.10
50	200	0.10	Steady-State	0.8	0.33	0.00
200	200	0.10	Posicional	-	0.32	0.03
200	50	0.10	Posicional	-	0.32	0.13
200	100	0.10	Steady-State	0.3	0.32	0.05
100	50	0.05	Steady-State	0.8	0.30	0.10
100	100	0.10	Steady-State	0.5	0.30	0.06
200	50	0.05	Steady-State	0.5	0.29	0.08
100	50	0.05	Posicional	-	0.29	0.22
50	200	0.05	Posicional	-	0.27	0.03
50	100	0.10	Steady-State	1.0	0.27	0.02
50	100	0.05	Steady-State	0.6	0.26	0.05
200	50	0.10	Steady-State	0.6	0.25	0.06
50	100	0.10	Steady-State	0.8	0.25	0.03
50	100	0.05	Steady-State	0.7	0.24	0.04
100	50	0.10	Steady-State	0.9	0.23	0.07
50	200	0.10	Steady-State	0.4	0.21	0.03

Continua ...

Tabela 10 – Calibragem do AG após as mudanças propostas. Taxa de sucesso dada após 100 execuções de cada conjunto de condições aplicado ao *experimento base*.

p_{ag}	g_{ag}	m_{ag}	Sobrevivência	s_{ag}	Histograma 10%	Melhor solução
50	200	0.05	Steady-State	0.3	0.21	0.06
100	50	0.05	Steady-State	0.7	0.20	0.06
100	100	0.10	Steady-State	0.4	0.20	0.02
50	100	0.10	Steady-State	0.9	0.20	0.04
200	200	0.05	Steady-State	0.1	0.19	0.06
100	200	0.10	Steady-State	0.2	0.19	0.05
100	100	0.05	Steady-State	0.3	0.18	0.04
200	50	0.10	Steady-State	0.5	0.16	0.04
200	50	0.05	Steady-State	0.4	0.16	0.03
200	200	0.10	Steady-State	0.1	0.16	0.02
200	100	0.05	Steady-State	0.2	0.15	0.02
100	200	0.10	Posicional	-	0.13	0.01
100	50	0.10	Steady-State	0.8	0.13	0.04
50	100	0.05	Steady-State	0.5	0.13	0.03
50	100	0.10	Steady-State	0.5	0.12	0.01
100	50	0.10	Steady-State	0.7	0.11	0.01
100	50	0.05	Steady-State	0.5	0.11	0.04
50	100	0.10	Posicional	-	0.11	0.04
200	100	0.10	Steady-State	0.2	0.10	0.00
100	50	0.05	Steady-State	0.6	0.10	0.05
50	50	0.05	Steady-State	1.0	0.10	0.01
50	100	0.10	Steady-State	0.7	0.10	0.01
100	100	0.10	Steady-State	0.3	0.09	0.02
50	200	0.10	Steady-State	0.3	0.09	0.00
50	100	0.10	Steady-State	0.6	0.08	0.02
200	50	0.10	Steady-State	0.4	0.07	0.00
200	50	0.10	Steady-State	0.3	0.07	0.00
50	50	0.05	Steady-State	0.8	0.07	0.06
50	100	0.05	Steady-State	0.4	0.07	0.02
200	200	0.05	Elitista	-	0.06	0.04
200	50	0.05	Steady-State	0.3	0.06	0.01
100	50	0.10	Posicional	-	0.06	0.04
100	50	0.10	Steady-State	0.6	0.06	0.03
50	50	0.05	Posicional	-	0.05	0.03
50	50	0.10	Steady-State	1.0	0.05	0.00
50	50	0.05	Steady-State	0.9	0.05	0.02
50	200	0.05	Steady-State	0.2	0.05	0.02
100	100	0.05	Steady-State	0.2	0.04	0.01
50	200	0.10	Steady-State	0.2	0.04	0.01
100	50	0.05	Steady-State	0.4	0.03	0.04
100	200	0.05	Steady-State	0.1	0.03	0.02

Continua ...

Tabela 10 – Calibragem do AG após as mudanças propostas. Taxa de sucesso dada após 100 execuções de cada conjunto de condições aplicado ao *experimento base*.

p_{ag}	g_{ag}	m_{ag}	Sobrevivência	s_{ag}	Histograma 10%	Melhor solução
50	200	0.10	Posicional	-	0.03	0.02
50	50	0.10	Posicional	-	0.03	0.01
50	200	0.05	Elitista	-	0.03	0.03
50	50	0.10	Steady-State	0.8	0.03	0.02
50	50	0.05	Steady-State	0.7	0.03	0.03
50	50	0.05	Steady-State	0.6	0.03	0.02
200	100	0.05	Geracional	-	0.02	0.01
200	100	0.05	Elitista	-	0.02	0.01
200	50	0.10	Elitista	-	0.02	0.00
200	100	0.05	Steady-State	0.1	0.02	0.00
100	200	0.05	Elitista	-	0.02	0.05
100	200	0.10	Steady-State	0.1	0.02	0.00
50	50	0.10	Steady-State	0.9	0.02	0.02
50	50	0.10	Steady-State	0.6	0.02	0.01
50	100	0.10	Steady-State	0.4	0.02	0.02
50	100	0.05	Steady-State	0.3	0.02	0.00
200	200	0.10	Elitista	-	0.01	0.01
200	50	0.10	Steady-State	0.2	0.01	0.00
200	100	0.10	Steady-State	0.1	0.01	0.00
100	100	0.05	Elitista	-	0.01	0.00
100	50	0.05	Elitista	-	0.01	0.00
50	50	0.10	Steady-State	0.7	0.01	0.01
50	100	0.10	Steady-State	0.3	0.01	0.00
50	100	0.05	Steady-State	0.2	0.01	0.00
200	200	0.10	Geracional	-	0.00	0.00
200	200	0.05	Geracional	-	0.00	0.01
200	100	0.10	Geracional	-	0.00	0.00
200	50	0.10	Geracional	-	0.00	0.00
200	50	0.05	Geracional	-	0.00	0.00
200	100	0.10	Elitista	-	0.00	0.00
200	50	0.05	Elitista	-	0.00	0.00
200	50	0.05	Steady-State	0.2	0.00	0.00
200	50	0.10	Steady-State	0.1	0.00	0.00
200	50	0.05	Steady-State	0.1	0.00	0.00
100	200	0.10	Geracional	-	0.00	0.00
100	200	0.05	Geracional	-	0.00	0.02
100	100	0.10	Geracional	-	0.00	0.00
100	100	0.05	Geracional	-	0.00	0.00
100	50	0.10	Geracional	-	0.00	0.00
100	50	0.05	Geracional	-	0.00	0.00
100	200	0.10	Elitista	-	0.00	0.01

Continua ...

Tabela 10 – Calibragem do AG após as mudanças propostas. Taxa de sucesso dada após 100 execuções de cada conjunto de condições aplicado ao *experimento base*.

p_{ag}	g_{ag}	m_{ag}	Sobrevivência	s_{ag}	Histograma 10%	Melhor solução
100	100	0.10	Elitista	-	0.00	0.00
100	50	0.10	Elitista	-	0.00	0.00
100	50	0.10	Steady-State	0.5	0.00	0.01
100	50	0.10	Steady-State	0.4	0.00	0.00
100	50	0.10	Steady-State	0.3	0.00	0.00
100	50	0.05	Steady-State	0.3	0.00	0.00
100	100	0.10	Steady-State	0.2	0.00	0.02
100	50	0.10	Steady-State	0.2	0.00	0.00
100	50	0.05	Steady-State	0.2	0.00	0.00
100	100	0.10	Steady-State	0.1	0.00	0.00
100	100	0.05	Steady-State	0.1	0.00	0.00
100	50	0.10	Steady-State	0.1	0.00	0.00
100	50	0.05	Steady-State	0.1	0.00	0.00
50	200	0.10	Geracional	-	0.00	0.00
50	200	0.05	Geracional	-	0.00	0.01
50	100	0.10	Geracional	-	0.00	0.00
50	100	0.05	Geracional	-	0.00	0.00
50	50	0.10	Geracional	-	0.00	0.00
50	50	0.05	Geracional	-	0.00	0.00
50	200	0.10	Elitista	-	0.00	0.00
50	100	0.10	Elitista	-	0.00	0.00
50	100	0.05	Elitista	-	0.00	0.00
50	50	0.10	Elitista	-	0.00	0.00
50	50	0.05	Elitista	-	0.00	0.00
50	50	0.10	Steady-State	0.5	0.00	0.00
50	50	0.05	Steady-State	0.5	0.00	0.01
50	50	0.10	Steady-State	0.4	0.00	0.00
50	50	0.05	Steady-State	0.4	0.00	0.00
50	50	0.10	Steady-State	0.3	0.00	0.00
50	50	0.05	Steady-State	0.3	0.00	0.00
50	100	0.10	Steady-State	0.2	0.00	0.00
50	50	0.10	Steady-State	0.2	0.00	0.00
50	50	0.05	Steady-State	0.2	0.00	0.00
50	200	0.10	Steady-State	0.1	0.00	0.00
50	200	0.05	Steady-State	0.1	0.00	0.00
50	100	0.10	Steady-State	0.1	0.00	0.00
50	100	0.05	Steady-State	0.1	0.00	0.00
50	50	0.10	Steady-State	0.1	0.00	0.00
50	50	0.05	Steady-State	0.1	0.00	0.00

Tabela 11 – Critérios que apresentaram menor mediana para a variável dependente *pico*. \bar{x} representa a média da amostra, s o desvio padrão, Md a mediana e iqr o intervalo interquartil. Percentual de suscetíveis representa a fração da população que não foi infectada ou vacinada no experimento.

Rede	pv	Critério	Pico				Post-hoc			% Suscetíveis
			Md	iqr	\bar{x}	s	p-valor	U-val	Equivalente	
ba400-m3	0.15	pagerank	79	17	79.26	12.36	1.00	524014.0	degree	28.05
ba400-m3	0.20	degree	46	14	47.11	9.76	1.00	482801.5	pagerank	45.61
ba400-m3	0.25	pagerank	32	9	32.45	6.01	0.83	526556.5	degree	57.80
ba400-m3	0.30	pagerank	27	6	27.41	4.08	0.46	529564.5	degree	58.05
ba400-m3	0.35	degree	24	5	24.76	3.42	0.76	526921.0	pagerank	54.88
ba400-m3	0.40	degree	22	4	23.08	2.78				51.46
ba400-m4	0.15	degree	140	15	140.03	11.66	1.00	495818.5	pagerank	10.98
ba400-m4	0.20	pagerank	100	17	99.54	12.64	1.00	524373.5	degree	17.07
ba400-m4	0.25	degree	67	15	66.34	11.19	1.00	481994.0	pagerank	26.83
ba400-m4	0.30	pagerank	43	10	43.25	8.46	0.20	533379.5	degree	42.68
ba400-m4	0.35	pagerank	31	8	31.22	5.71				48.54
ba400-m4	0.40	degree	26	6	26.42	4.07	1.00	521881.0	pagerank	48.29
ba400-m5	0.15	pagerank	182	15	181.82	10.94	0.48	529419.5	degree	4.39
ba400-m5	0.15	pagerank	182	15	181.82	10.94	0.31	531510.0	gasir	4.39
ba400-m5	0.20	pagerank	142	16	141.97	11.50				7.80
ba400-m5	0.25	pagerank	105	15	105.15	11.63				13.17
ba400-m5	0.30	pagerank	73	15	72.53	11.03	1.00	523437.5	degree	20.73
ba400-m5	0.35	pagerank	46	12	46.09	8.87				33.90
ba400-m5	0.40	pagerank	33	9	33.53	6.37				39.02
ba400-m6	0.15	gasir	208	14	208.17	9.79	1.00	495060.5	pagerank	3.41
ba400-m6	0.15	gasir	208	14	208.17	9.79	1.00	521002.5	betweenness	3.41
ba400-m6	0.15	gasir	208	14	208.17	9.79	1.00	510197.0	degree	3.41
ba400-m6	0.20	pagerank	170	14	170.72	10.76	1.00	506447.5	degree	5.61

Continua ...

Tabela 11 – Critérios que apresentaram menor mediana para a variável dependente *pico*. \bar{x} representa a média da amostra, s o desvio padrão, Md a mediana e iqr o intervalo interquartil. Percentual de suscetíveis representa a fração da população que não foi infectada no experimento.

Rede	pv	Critério	Pico				Post-hoc			% Suscetíveis
			Md	iqr	\bar{x}	s	p-valor	U-val	Equivalente	
ba400-m6	0.25	pagerank	139	14	138.42	10.53	0.12	535801.0	betweenness	9.02
ba400-m6	0.30	pagerank	101	15	100.68	10.61				14.63
ba400-m6	0.35	degree	68	14	67.63	10.59	1.00	491983.0	pagerank	21.46
ba400-m6	0.40	pagerank	38	11	38.67	7.94				34.39
moreno health	0.15	pagerank	1254	37	1253.96	28.18				2.80
moreno health	0.20	pagerank	1076	39	1075.34	28.85				4.14
moreno health	0.25	pagerank	910	37	908.00	28.48				5.51
moreno health	0.30	pagerank	722	42	720.63	29.88				7.72
moreno health	0.35	pagerank	556	41	556.14	29.14				11.19
moreno health	0.40	pagerank	409	35	408.82	26.50				15.56
moreno highschool	0.15	degree	34	6	34.07	4.27	0.70	527433.5	eigenvector	1.43
moreno highschool	0.15	degree	34	6	34.07	4.27	1.00	523830.5	gasir	1.43
moreno highschool	0.15	degree	34	6	34.07	4.27	1.00	507323.0	pagerank	1.43
moreno highschool	0.20	pagerank	27	6	27.27	4.49				2.86
moreno highschool	0.25	betweenness	24	6	23.65	5.43	1.00	489980.5	degree	15.71
moreno highschool	0.25	betweenness	24	6	23.65	5.43	1.00	504003.5	gasir	15.71
moreno highschool	0.25	betweenness	24	6	23.65	5.43	1.00	504797.0	pagerank	15.71
moreno highschool	0.30	gasir	15	7	15.61	5.30	1.00	514963.5	betweenness	25.71
moreno highschool	0.35	degree	12	5	12.47	4.05	1.00	520778.0	betweenness	28.57
moreno highschool	0.40	betweenness	8	7	9.35	4.09	1.00	474507.5	pagerank	40.00
moreno innovation	0.15	gasir	97	21	97.28	14.24	1.00	503551.5	pagerank	4.15
moreno innovation	0.15	gasir	97	21	97.28	14.24	1.00	524481.0	degree	4.15
moreno innovation	0.20	degree	73	12	72.52	9.10	1.00	480317.5	pagerank	21.58

Continua ...

Tabela 11 – Critérios que apresentaram menor mediana para a variável dependente *pico*. \bar{x} representa a média da amostra, s o desvio padrão, Md a mediana e iqr o intervalo interquartil. Percentual de suscetíveis representa a fração da população que não foi infectada no experimento.

Rede	pv	Critério	Pico				Post-hoc			% Suscetíveis
			Md	iqr	\bar{x}	s	p-valor	U-val	Equivalente	
moreno innovation	0.25	gasir	57	13	57.06	9.64				10.37
moreno innovation	0.30	degree	45	11	45.13	8.03	1.00	493811.5	gasir	27.39
moreno innovation	0.30	degree	45	11	45.13	8.03	0.55	528689.5	pagerank	27.39
moreno innovation	0.35	pagerank	28	10	28.60	7.07				33.20
moreno innovation	0.40	pagerank	19	6	19.79	4.62				41.49
moreno oz	0.15	gasir	138	9	137.45	6.77				0.00
moreno oz	0.20	gasir	126	9	126.04	6.36	0.21	533184.0	betweenness	0.46
moreno oz	0.25	gasir	114	9	113.47	6.64				0.46
moreno oz	0.30	gasir	101	8	101.07	6.43				0.92
moreno oz	0.35	betweenness	91	8	91.02	6.37	0.42	530029.5	gasir	0.92
moreno oz	0.35	betweenness	91	8	91.02	6.37	1.00	475147.0	pagerank	0.92
moreno oz	0.40	gasir	77	9	76.39	6.50				3.23

Tabela 12 – Critérios que apresentaram menor mediana para a variável dependente β . \bar{x} representa a média da amostra, s o desvio padrão, Md a mediana e iqr o intervalo interquartil. Percentual de suscetíveis representa a fração da população que não foi infectada ou vacinada no experimento.

Rede	pv	Critério	Beta				Post-hoc			% Suscetíveis
			Md	iqr	\bar{x}	s	p-valor	U-val	Equivalente	
ba400-m3	0.15	degree	0.35	0.09	0.35	0.07	0.88	526304.0	pagerank	27.07
ba400-m3	0.20	degree	0.18	0.03	0.18	0.04	1.00	479944.0	pagerank	45.61
ba400-m3	0.25	degree	0.13	0.02	0.13	0.02				56.83
ba400-m3	0.30	pagerank	0.10	0.02	0.10	0.02				58.05
ba400-m3	0.35	pagerank	0.08	0.03	0.08	0.02				55.37
ba400-m3	0.40	degree	0.07	0.03	0.07	0.03				51.46
ba400-m4	0.15	degree	0.61	0.08	0.61	0.06	1.00	490136.0	pagerank	10.98
ba400-m4	0.20	pagerank	0.46	0.08	0.46	0.06	1.00	520806.0	degree	17.07
ba400-m4	0.25	degree	0.31	0.10	0.31	0.07	1.00	496416.0	pagerank	26.83
ba400-m4	0.30	degree	0.17	0.04	0.18	0.03				40.49
ba400-m4	0.35	pagerank	0.13	0.03	0.13	0.02				48.54
ba400-m4	0.40	degree	0.10	0.03	0.10	0.02	0.10	536672.0	pagerank	48.29
ba400-m5	0.15	degree	0.80	0.09	0.80	0.06	0.41	469848.0	gasir	4.15
ba400-m5	0.15	degree	0.80	0.09	0.80	0.06	1.00	500730.0	pagerank	4.15
ba400-m5	0.20	pagerank	0.64	0.08	0.64	0.06				7.80
ba400-m5	0.25	pagerank	0.51	0.08	0.51	0.06				13.17
ba400-m5	0.30	degree	0.38	0.10	0.38	0.07	1.00	514737.0	pagerank	20.49
ba400-m5	0.35	pagerank	0.19	0.05	0.20	0.05				33.90
ba400-m5	0.40	pagerank	0.15	0.03	0.15	0.03				39.02
ba400-m6	0.15	degree	1.00	0.10	1.01	0.07	1.00	518478.0	gasir	3.17
ba400-m6	0.15	degree	1.00	0.10	1.01	0.07	1.00	525312.0	pagerank	3.17
ba400-m6	0.15	degree	1.00	0.10	1.01	0.07	1.00	509820.0	betweenness	3.17
ba400-m6	0.20	degree	0.83	0.09	0.84	0.07	0.60	528234.0	pagerank	5.85

Continua ...

Tabela 12 – Critérios que apresentaram menor mediana para a variável dependente β . \bar{x} representa a média da amostra, s o desvio padrão, Md a mediana e iqr o intervalo interquartil. Percentual de suscetíveis representa a fração da população que não foi infectada no experimento.

Rede	pv	Critério	Beta				Post-hoc			% Suscetíveis
			Md	iqr	\bar{x}	s	p-valor	U-val	Equivalente	
ba400-m6	0.25	betweenness	0.69	0.09	0.70	0.07	0.20	466452.0	degree	8.05
ba400-m6	0.25	betweenness	0.69	0.09	0.70	0.07	1.00	498231.0	pagerank	8.05
ba400-m6	0.30	pagerank	0.54	0.09	0.54	0.07				14.63
ba400-m6	0.35	degree	0.36	0.11	0.36	0.08	1.00	480237.0	pagerank	21.46
ba400-m6	0.40	pagerank	0.17	0.04	0.18	0.04				34.39
moreno health	0.15	pagerank	0.96	0.04	0.96	0.03				2.80
moreno health	0.20	pagerank	0.84	0.04	0.84	0.03				4.14
moreno health	0.25	pagerank	0.73	0.04	0.73	0.03				5.51
moreno health	0.30	pagerank	0.61	0.04	0.61	0.03				7.72
moreno health	0.35	pagerank	0.51	0.04	0.51	0.03				11.19
moreno health	0.40	pagerank	0.40	0.04	0.40	0.03				15.56
moreno highschool	0.15	gasir	0.88	0.22	0.89	0.17	0.41	469853.0	pagerank	1.43
moreno highschool	0.15	gasir	0.88	0.22	0.89	0.17	1.00	521096.0	degree	1.43
moreno highschool	0.20	pagerank	0.71	0.22	0.72	0.16				2.86
moreno highschool	0.25	gasir	0.65	0.24	0.67	0.18	1.00	516563.0	pagerank	5.71
moreno highschool	0.25	gasir	0.65	0.24	0.67	0.18	1.00	509067.0	degree	5.71
moreno highschool	0.30	betweenness	0.45	0.33	0.44	0.23	1.00	504032.0	gasir	20.00
moreno highschool	0.35	degree	0.28	0.33	0.35	0.23	1.00	494943.0	betweenness	28.57
moreno highschool	0.40	betweenness	0.16	0.24	0.21	0.78	1.00	490417.0	pagerank	40.00
moreno innovation	0.15	degree	0.81	0.15	0.80	0.13	1.00	499170.0	gasir	3.73
moreno innovation	0.15	degree	0.81	0.15	0.80	0.13	1.00	507948.0	pagerank	3.73
moreno innovation	0.20	degree	0.66	0.17	0.66	0.15	0.23	467212.0	pagerank	21.58
moreno innovation	0.25	gasir	0.46	0.16	0.46	0.12				10.37

Continua ...

Tabela 12 – Critérios que apresentaram menor mediana para a variável dependente *beta*. \bar{x} representa a média da amostra, *s* o desvio padrão, *Md* a mediana e *iqr* o intervalo interquartil. Percentual de suscetíveis representa a fração da população que não foi infectada no experimento.

Rede	<i>pv</i>	Critério	Beta				Post-hoc			% Suscetíveis
			<i>Md</i>	<i>iqr</i>	\bar{x}	<i>s</i>	p-valor	U-val	Equivalente	
moreno innovation	0.30	degree	0.34	0.23	0.35	0.14	1.00	490654.0	gasir	27.39
moreno innovation	0.30	degree	0.34	0.23	0.35	0.14	1.00	510516.0	pagerank	27.39
moreno innovation	0.35	gasir	0.19	0.10	0.23	0.09	0.85	526455.0	pagerank	33.20
moreno innovation	0.35	gasir	0.19	0.10	0.23	0.09	1.00	493077.0	degree	33.20
moreno innovation	0.40	pagerank	0.14	0.04	0.14	0.04				41.49
moreno oz	0.15	betweenness	1.66	0.18	1.67	0.13	1.00	484733.0	degree	0.00
moreno oz	0.15	betweenness	1.66	0.18	1.67	0.13	1.00	512464.0	gasir	0.00
moreno oz	0.15	betweenness	1.66	0.18	1.67	0.13	1.00	488002.0	pagerank	0.00
moreno oz	0.20	gasir	1.53	0.18	1.53	0.13	1.00	496735.0	pagerank	0.46
moreno oz	0.20	gasir	1.53	0.18	1.53	0.13	0.28	531936.0	betweenness	0.46
moreno oz	0.20	gasir	1.53	0.18	1.53	0.13	0.18	533995.0	degree	0.46
moreno oz	0.25	gasir	1.39	0.18	1.40	0.14	1.00	480471.0	pagerank	0.46
moreno oz	0.30	pagerank	1.26	0.17	1.27	0.13	1.00	511904.0	gasir	0.46
moreno oz	0.35	gasir	1.16	0.18	1.16	0.13	1.00	485433.0	pagerank	0.92
moreno oz	0.35	gasir	1.16	0.18	1.16	0.13	1.00	516479.0	betweenness	0.92
moreno oz	0.40	pagerank	1.00	0.17	1.00	0.12	0.08	537468.0	betweenness	1.38

Tabela 13 – Critérios que apresentaram menor mediana para a variável dependente *afetados*. \bar{x} representa a média da amostra, s o desvio padrão, Md a mediana e iqr o intervalo interquartil. Percentual de suscetíveis representa a fração da população que não foi infectada ou vacinada no experimento.

Rede	pv	Critério	Afetados				Post-hoc			% Suscetíveis
			Md	iqr	\bar{x}	s	p-valor	U-val	Equivalente	
ba400-m3	0.15	pagerank	295	22	293.11	17.91				28.05
ba400-m3	0.20	degree	223	33	220.90	24.69	0.65	472134.5	pagerank	45.61
ba400-m3	0.25	pagerank	173	20	173.58	14.10				57.80
ba400-m3	0.30	pagerank	172	11	172.90	7.98				58.05
ba400-m3	0.35	pagerank	183	8	183.16	5.84				55.37
ba400-m3	0.40	pagerank	198	6	197.89	4.39				51.71
ba400-m4	0.15	pagerank	364	11	364.23	7.78	1.00	500893.0	degree	11.22
ba400-m4	0.20	degree	340	15	339.26	11.09	1.00	504432.0	pagerank	17.07
ba400-m4	0.25	pagerank	295	20	293.27	16.75				27.93
ba400-m4	0.30	pagerank	235	26	234.93	19.95				42.68
ba400-m4	0.35	pagerank	211	18	211.38	13.15				48.54
ba400-m4	0.40	pagerank	210	10	210.48	7.30				48.78
ba400-m5	0.15	gasir	391	6	391.02	4.07				4.63
ba400-m5	0.20	pagerank	378	8	377.77	6.02	0.07	537611.0	gasir	7.80
ba400-m5	0.25	degree	356	12	355.63	8.52	0.68	527647.5	pagerank	13.17
ba400-m5	0.30	pagerank	325	18	324.30	13.58	1.00	514957.0	degree	20.73
ba400-m5	0.35	pagerank	271	26	269.84	20.39				33.90
ba400-m5	0.40	pagerank	250	23	248.84	16.81				39.02
ba400-m6	0.15	gasir	396	5	395.43	3.02				3.41
ba400-m6	0.20	degree	386	6	385.64	4.08				5.85
ba400-m6	0.25	pagerank	373	7	373.04	5.44	1.00	513538.0	gasir	9.02
ba400-m6	0.30	pagerank	350	11	349.38	7.90				14.63
ba400-m6	0.35	degree	322	16	320.60	12.11	1.00	496240.0	pagerank	21.46

Continua ...

Tabela 13 – Critérios que apresentaram menor mediana para a variável dependente *afetados*. \bar{x} representa a média da amostra, s o desvio padrão, Md a mediana e iqr o intervalo interquartil. Percentual de suscetíveis representa a fração da população que não foi infectada no experimento.

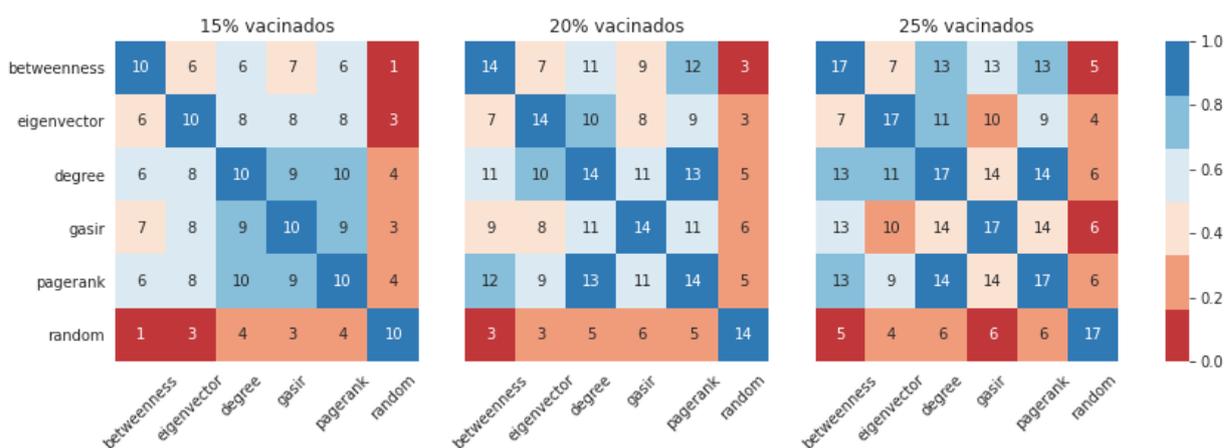
Rede	pv	Critério	Afetados				Post-hoc			% Suscetíveis
			Md	iqr	\bar{x}	s	p-valor	U-val	Equivalente	
ba400-m6	0.40	pagerank	269	23	267.46	18.47				34.39
moreno health	0.15	betweenness	2461	11	2460.70	8.27				3.07
moreno health	0.20	pagerank	2434	12	2433.78	8.93	0.43	529945.5	betweenness	4.14
moreno health	0.25	pagerank	2399	15	2398.58	11.40				5.51
moreno health	0.30	pagerank	2343	21	2342.29	14.57				7.72
moreno health	0.35	pagerank	2255	28	2254.72	20.01				11.19
moreno health	0.40	pagerank	2144	36	2142.45	27.19				15.56
moreno highschool	0.15	betweenness	69	2	68.70	1.32	0.11	465457.5	degree	1.43
moreno highschool	0.15	betweenness	69	2	68.70	1.32	0.14	466400.0	gasir	1.43
moreno highschool	0.15	betweenness	69	2	68.70	1.32	1.00	480821.0	pagerank	1.43
moreno highschool	0.20	betweenness	67	2	66.59	3.05				4.29
moreno highschool	0.25	betweenness	59	10	60.50	7.60				15.71
moreno highschool	0.30	gasir	52	18	53.03	10.68	0.15	534713.5	betweenness	25.71
moreno highschool	0.35	degree	50	9	49.03	7.87				28.57
moreno highschool	0.40	betweenness	42	16	45.04	8.76	1.00	510993.5	pagerank	40.00
moreno innovation	0.15	betweenness	209	43	213.01	25.47	1.00	490248.0	degree	13.28
moreno innovation	0.15	betweenness	209	43	213.01	25.47	1.00	475065.5	eigenvector	13.28
moreno innovation	0.15	betweenness	209	43	213.01	25.47	1.00	510118.0	gasir	13.28
moreno innovation	0.15	betweenness	209	43	213.01	25.47	1.00	497387.0	pagerank	13.28
moreno innovation	0.20	degree	189	42	200.71	21.88	1.00	488568.5	pagerank	21.58
moreno innovation	0.25	degree	186	37	194.09	21.43	0.87	473699.5	betweenness	22.82
moreno innovation	0.30	degree	175	36	180.74	21.30	1.00	475727.0	gasir	27.39
moreno innovation	0.30	degree	175	36	180.74	21.30	0.15	465362.0	pagerank	27.39

Continua ...

Tabela 13 – Critérios que apresentaram menor mediana para a variável dependente *afetados*. \bar{x} representa a média da amostra, s o desvio padrão, Md a mediana e iqr o intervalo interquartil. Percentual de suscetíveis representa a fração da população que não foi infectada no experimento.

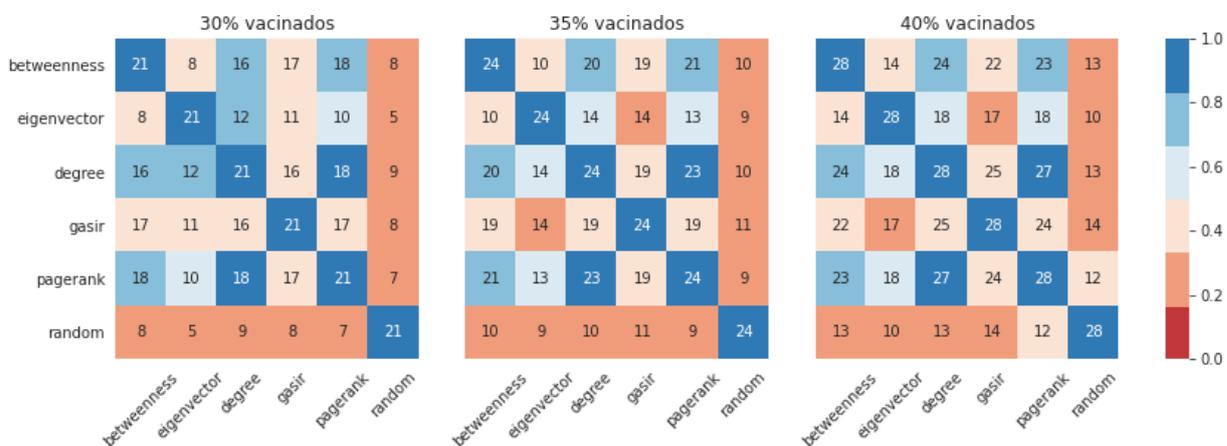
Rede	pv	Critério	Afetados				Post-hoc			% Suscetíveis
			Md	iqr	\bar{x}	s	p-valor	U-val	Equivalente	
moreno innovation	0.35	gasir	161	25	165.27	21.14	1.00	520693.0	pagerank	33.20
moreno innovation	0.35	gasir	161	25	165.27	21.14	1.00	499075.5	degree	33.20
moreno innovation	0.40	pagerank	141	19	143.20	15.06				41.49
moreno oz	0.15	eigenvector	216	0	215.96	0.43				0.46
moreno oz	0.20	betweenness	216	1	216.36	0.75				0.46
moreno oz	0.25	betweenness	216	1	215.63	1.08	1.00	517710.0	eigenvector	0.46
moreno oz	0.30	gasir	215	1	214.47	1.31				0.92
moreno oz	0.35	betweenness	215	2	214.76	1.50	0.06	462770.5	eigenvector	0.92
moreno oz	0.40	gasir	210	3	209.52	2.79				3.23

APÊNDICE B – Correlação entre as heurísticas de vacinação



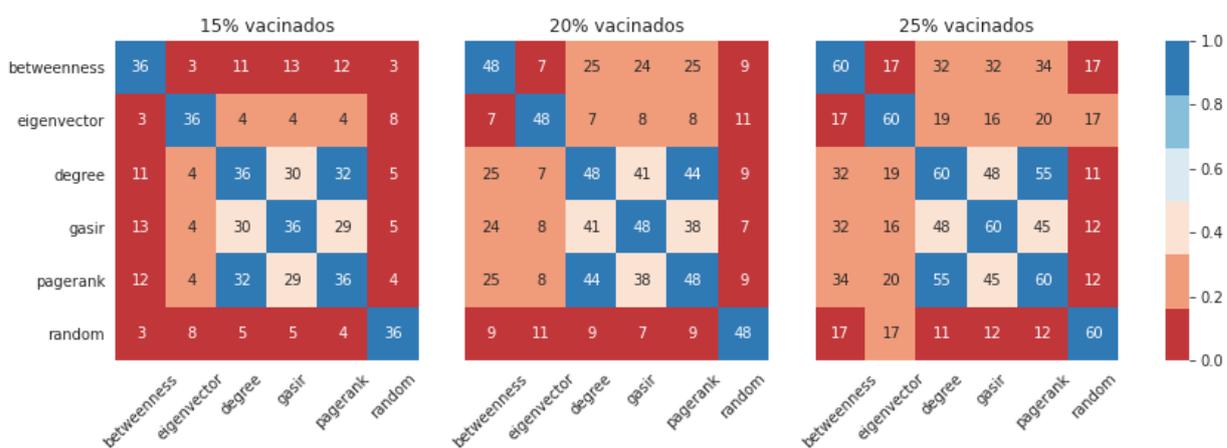
Fonte: O autor

Figura 28 – Similaridade e intersecção dos critérios de vacinação entre 15% e 25% na rede *highschool*



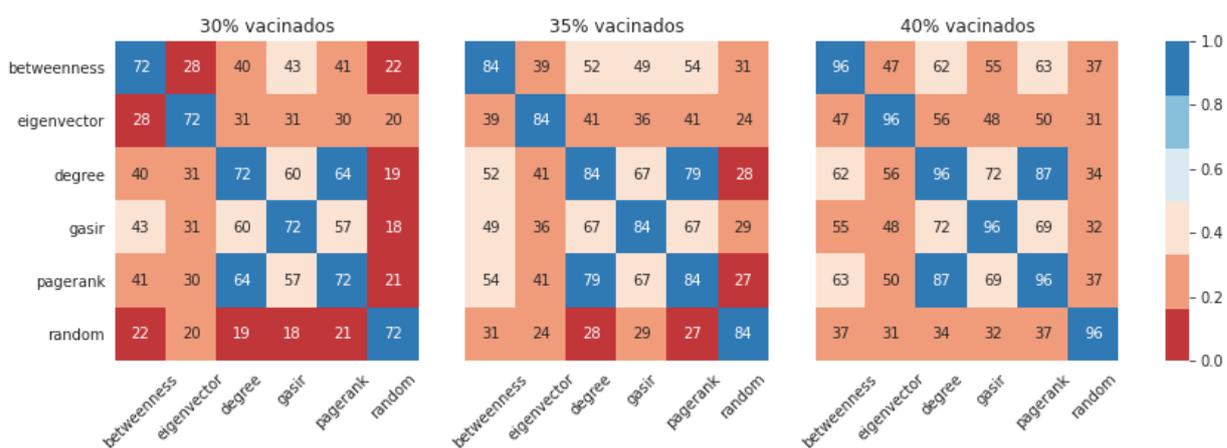
Fonte: O autor

Figura 29 – Similaridade e intersecção dos critérios de vacinação entre 30% e 40% na rede *highschool*



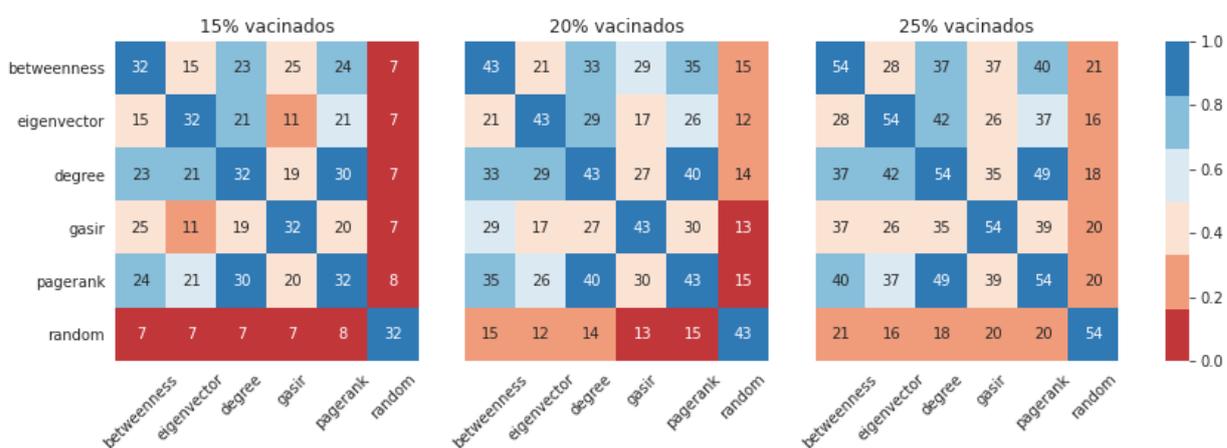
Fonte: O autor

Figura 30 – Similaridade e intersecção dos critérios de vacinação entre 15% e 25% na rede *innovation*



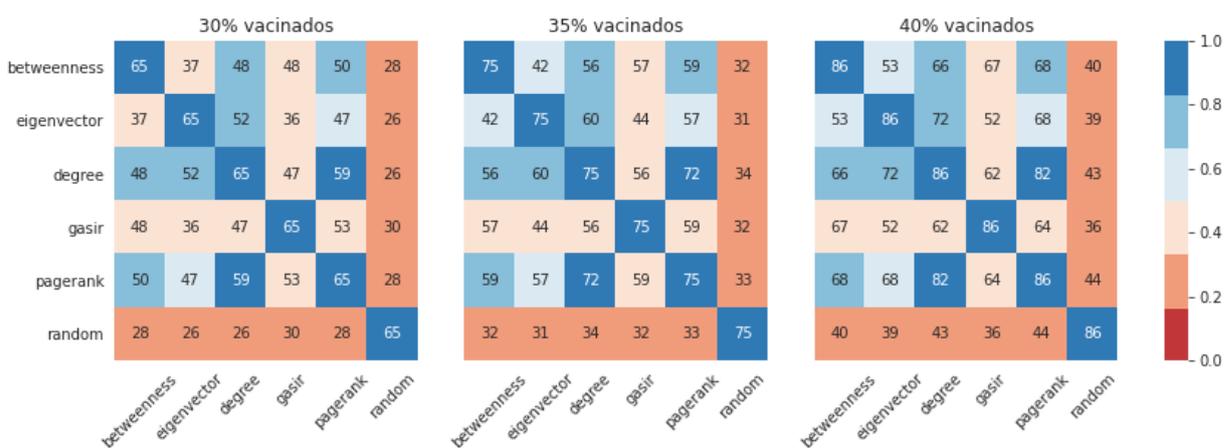
Fonte: O autor

Figura 31 – Similaridade e intersecção dos critérios de vacinação entre 30% e 40% na rede *innovation*



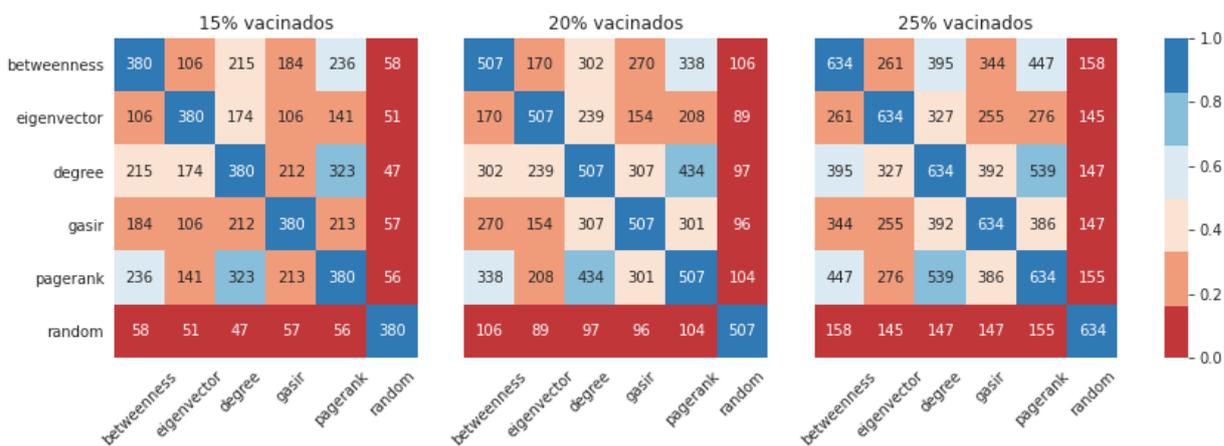
Fonte: O autor

Figura 32 – Similaridade e intersecção dos critérios de vacinação entre 15% e 25% na rede *oz*



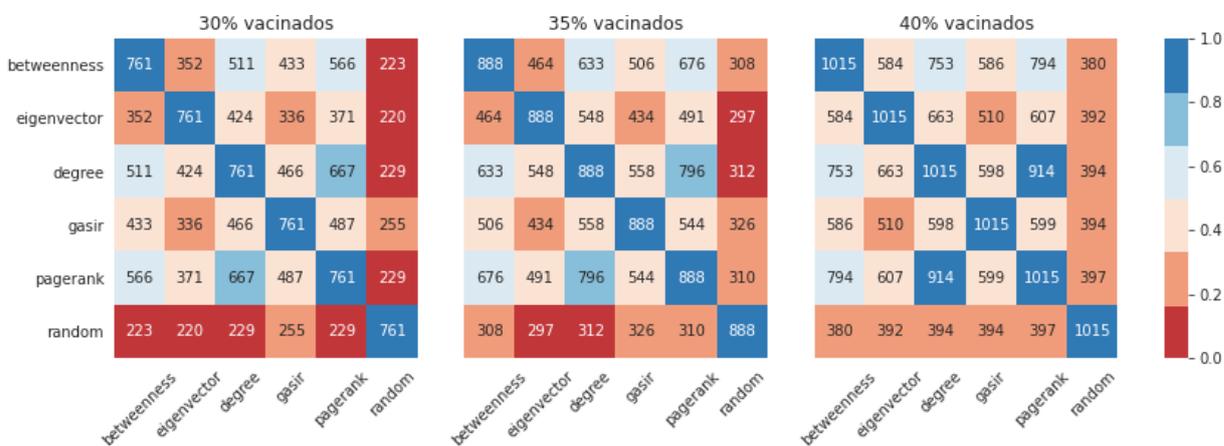
Fonte: O autor

Figura 33 – Similaridade e intersecção dos critérios de vacinação entre 30% e 40% na rede *oz*



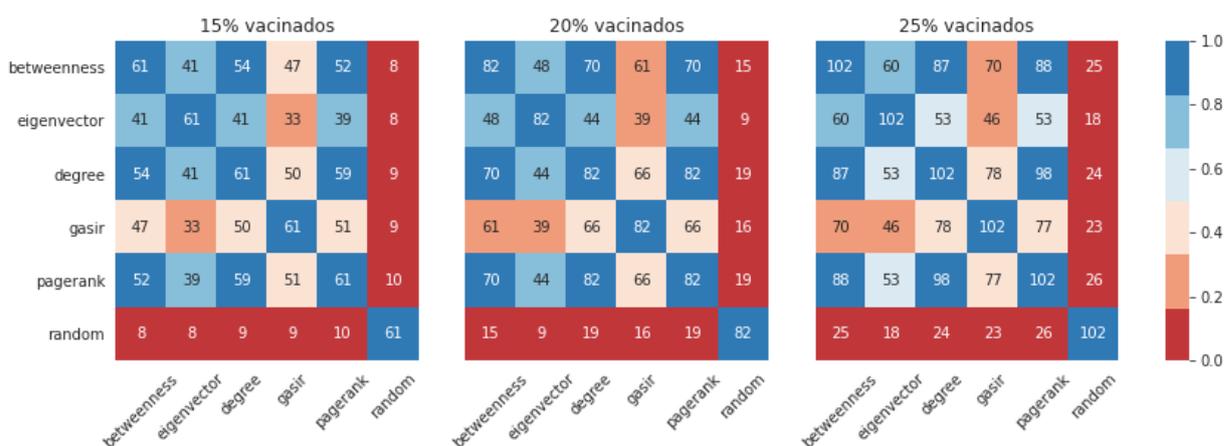
Fonte: O autor

Figura 34 – Similaridade e intersecção dos critérios de vacinação entre 15% e 25% na rede *health*



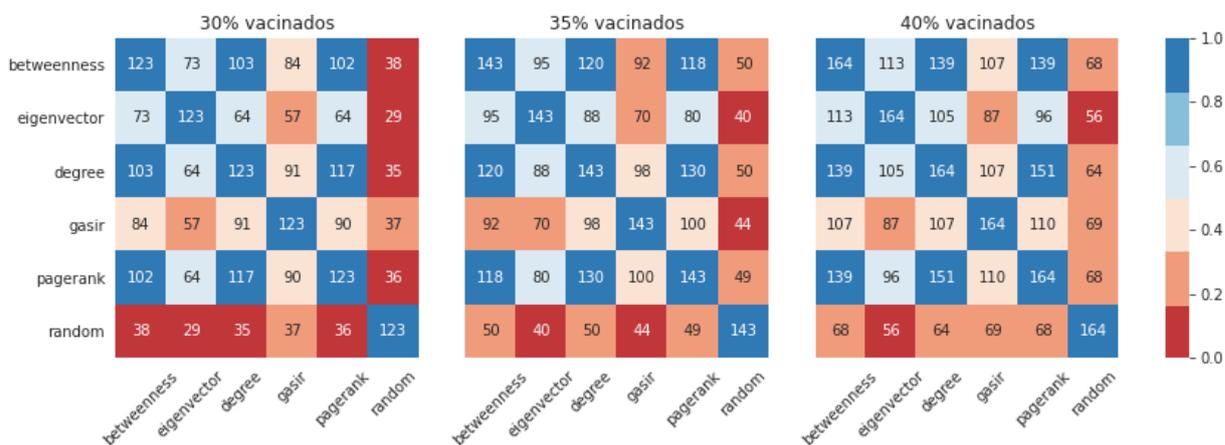
Fonte: O autor

Figura 35 – Similaridade e intersecção dos critérios de vacinação entre 30% e 40% na rede *health*



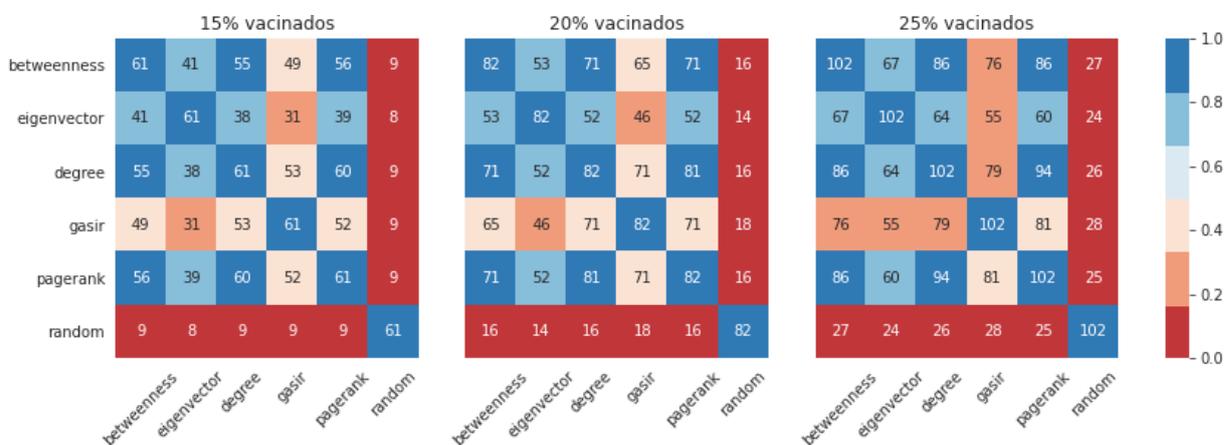
Fonte: O autor

Figura 36 – Similaridade e intersecção dos critérios de vacinação entre 15% e 25% na rede *ba400-m3*



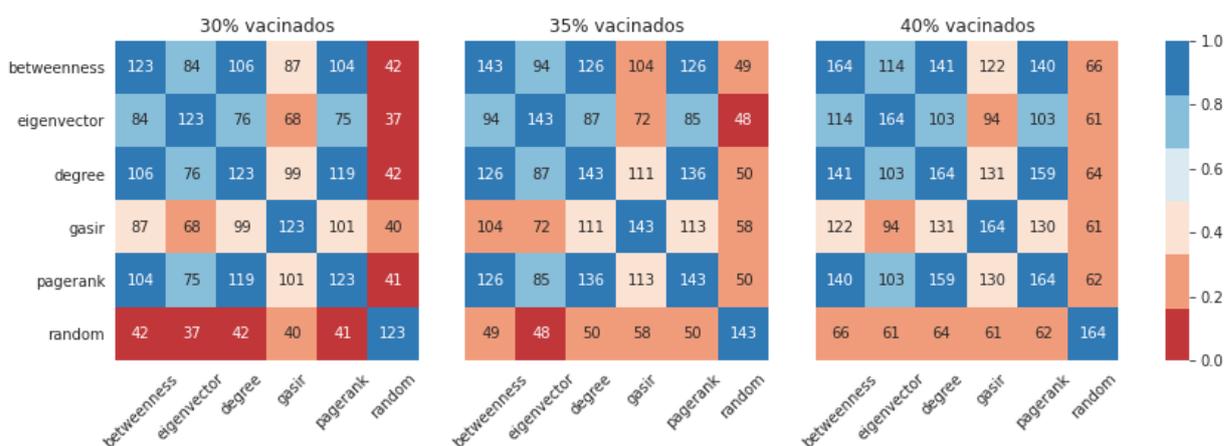
Fonte: O autor

Figura 37 – Similaridade e intersecção dos critérios de vacinação entre 30% e 40% na rede *ba400-m3*



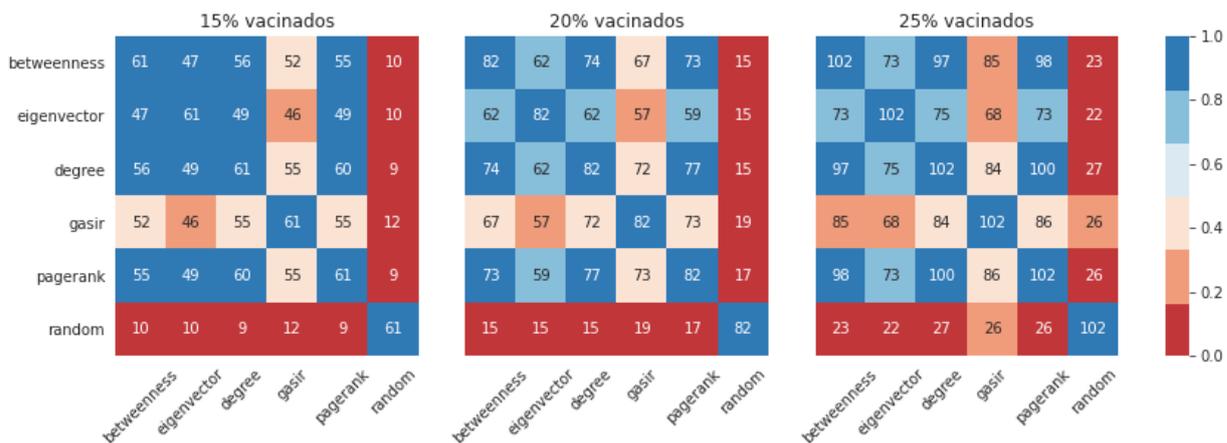
Fonte: O autor

Figura 38 – Similaridade e intersecção dos critérios de vacinação entre 15% e 25% na rede *ba400-m4*



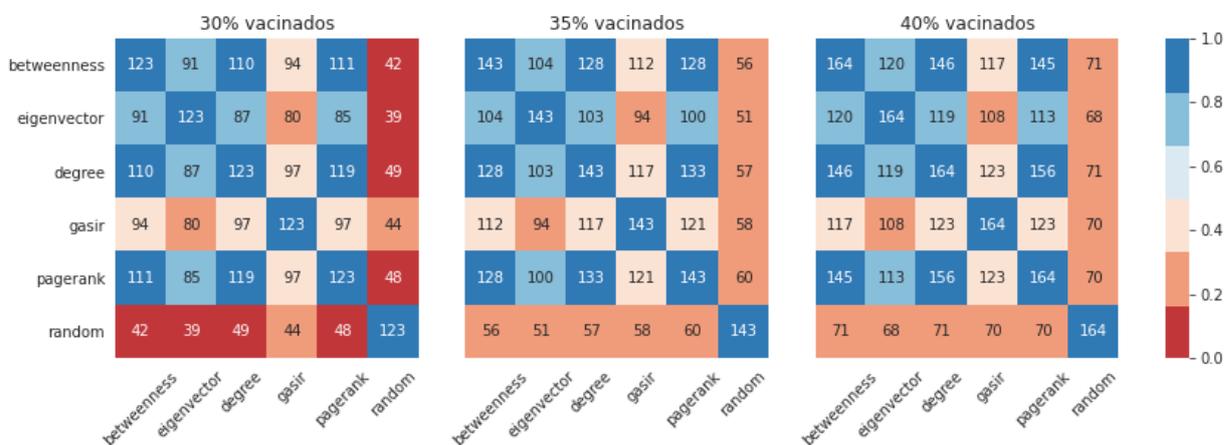
Fonte: O autor

Figura 39 – Similaridade e intersecção dos critérios de vacinação entre 30% e 40% na rede *ba.400-m4*



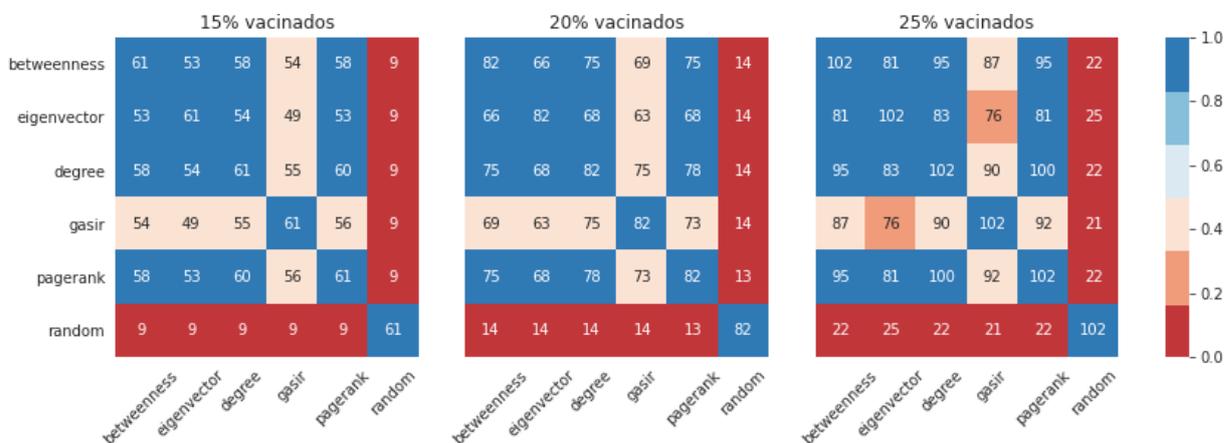
Fonte: O autor

Figura 40 – Similaridade e intersecção dos critérios de vacinação entre 15% e 25% na rede *ba.400-m5*



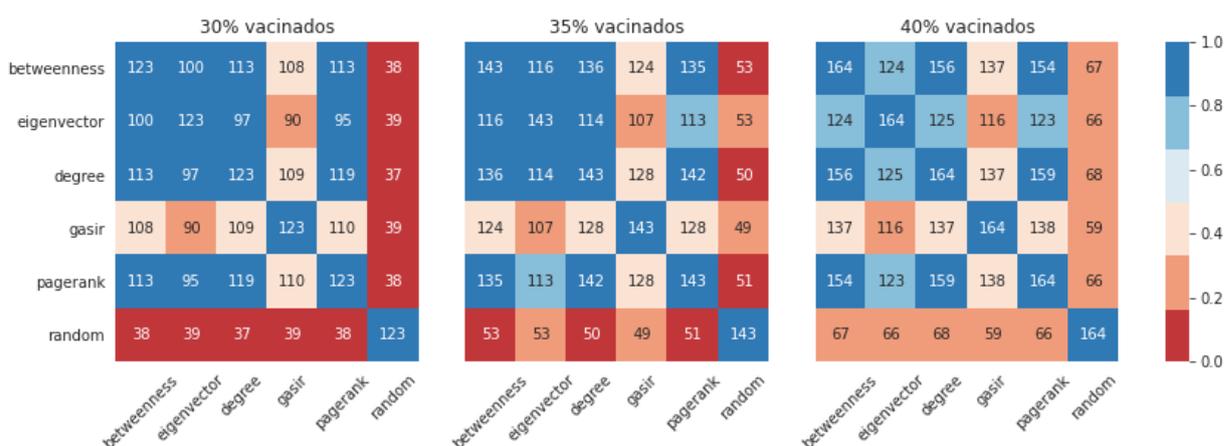
Fonte: O autor

Figura 41 – Similaridade e intersecção dos critérios de vacinação entre 30% e 40% na rede *ba.400-m5*



Fonte: O autor

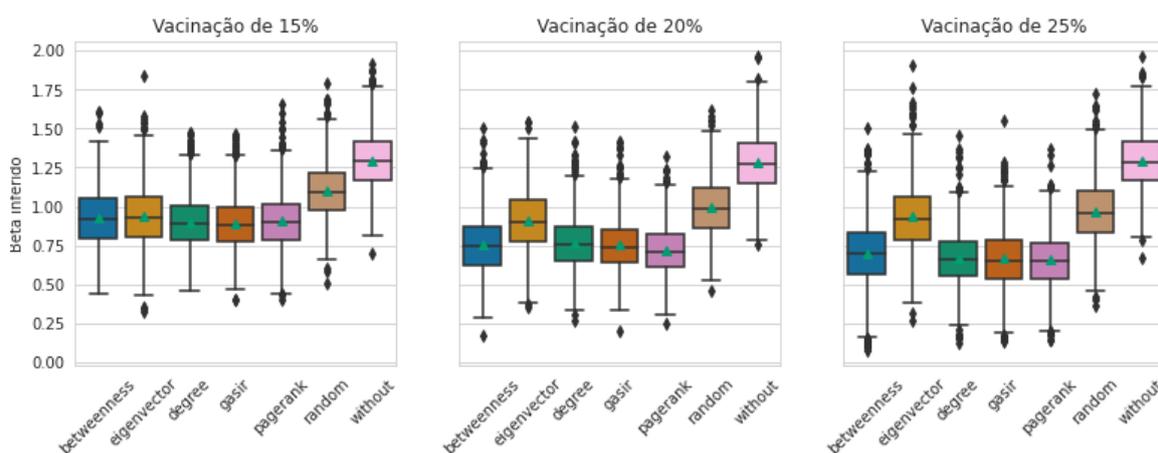
Figura 42 – Similaridade e intersecção dos critérios de vacinação entre 15% e 25% na rede *ba400-m6*



Fonte: O autor

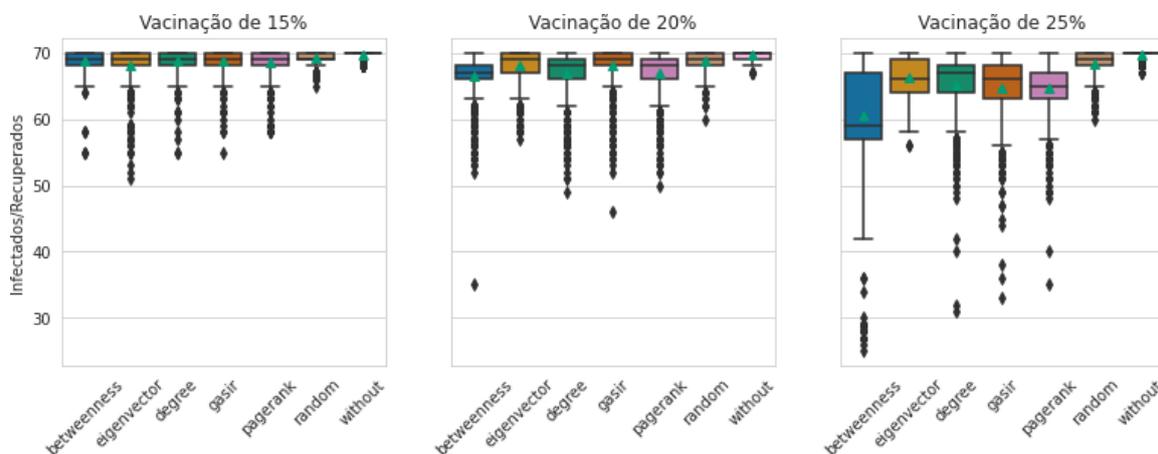
Figura 43 – Similaridade e intersecção dos critérios de vacinação entre 30% e 40% na rede *ba400-m6*

APÊNDICE C – Boxplot das variáveis dependentes consideradas no estudo.



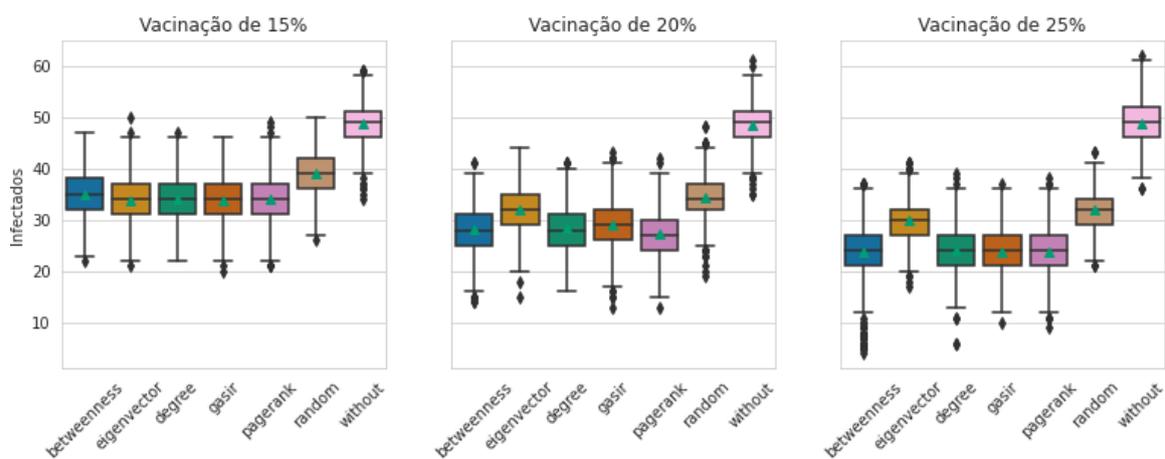
Fonte: O autor

Figura 44 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 15% e 25% na rede *highschool*.



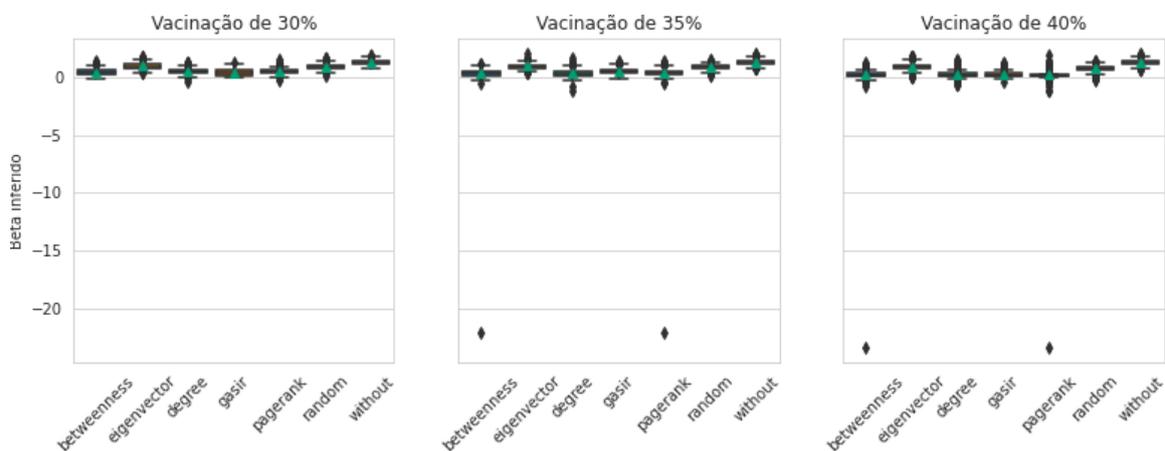
Fonte: O autor

Figura 45 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 15% e 25% na rede *highschool*.



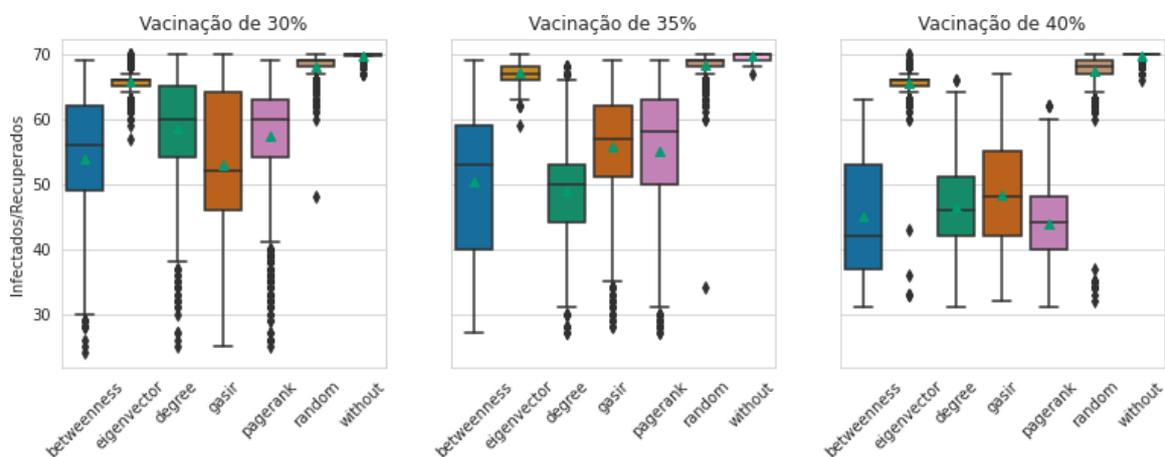
Fonte: O autor

Figura 46 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 15% e 25% na rede *highschool*.



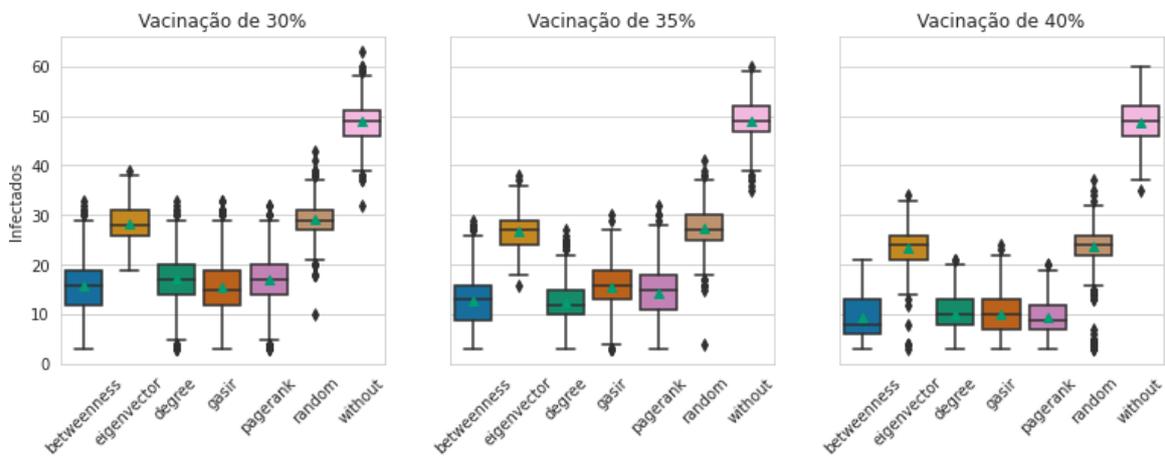
Fonte: O autor

Figura 47 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 30% e 40% na rede *highschool*.



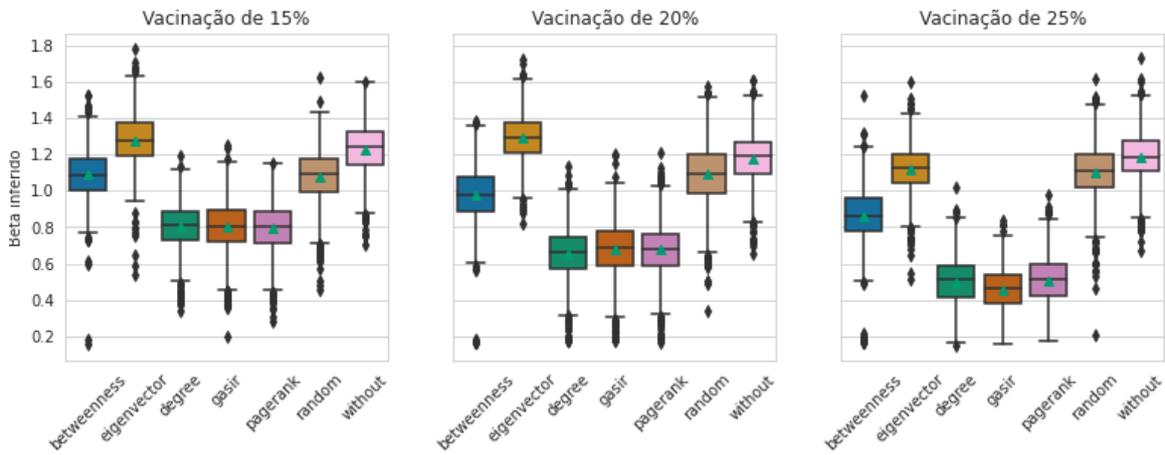
Fonte: O autor

Figura 48 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 30% e 40% na rede *highschool*.



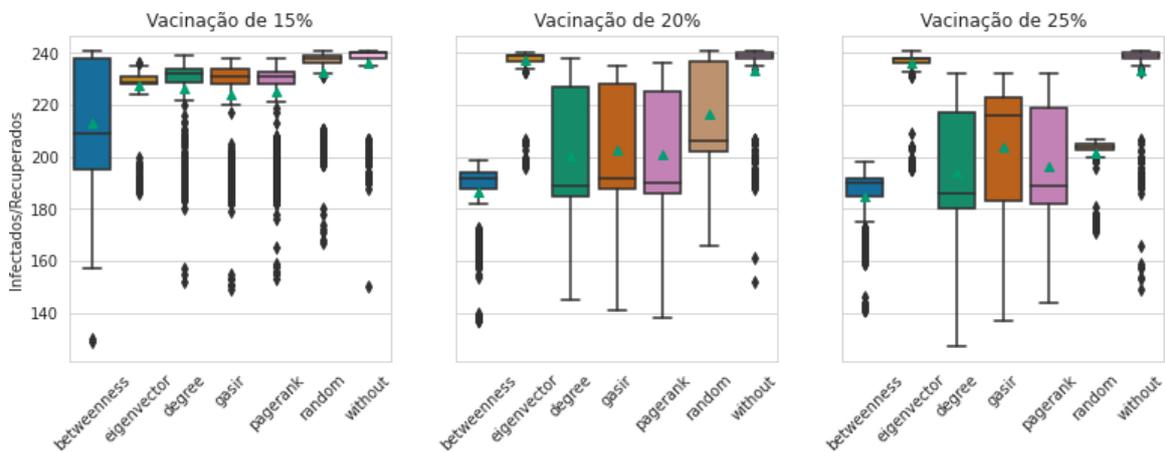
Fonte: O autor

Figura 49 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 30% e 40% na rede *highschool*.



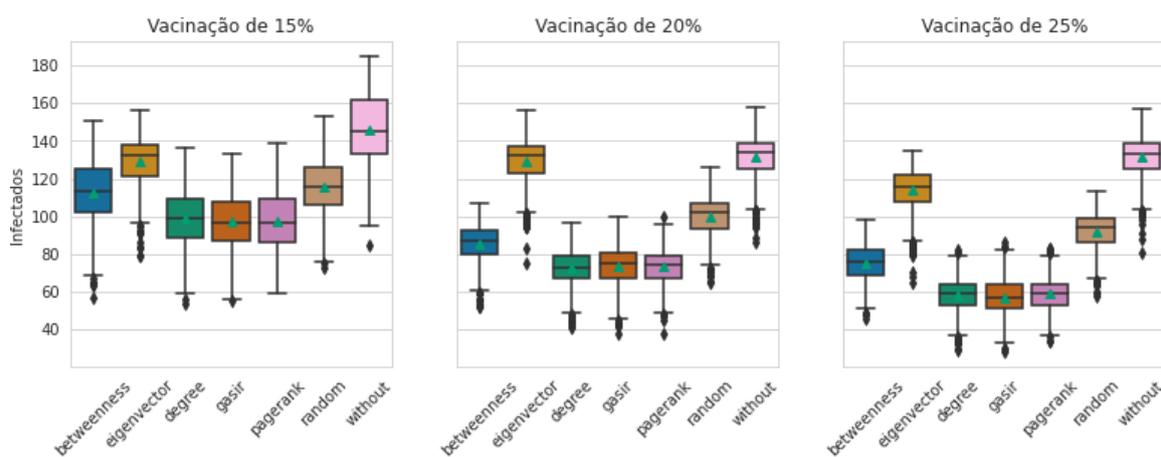
Fonte: O autor

Figura 50 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 15% e 25% na rede *innovation*.



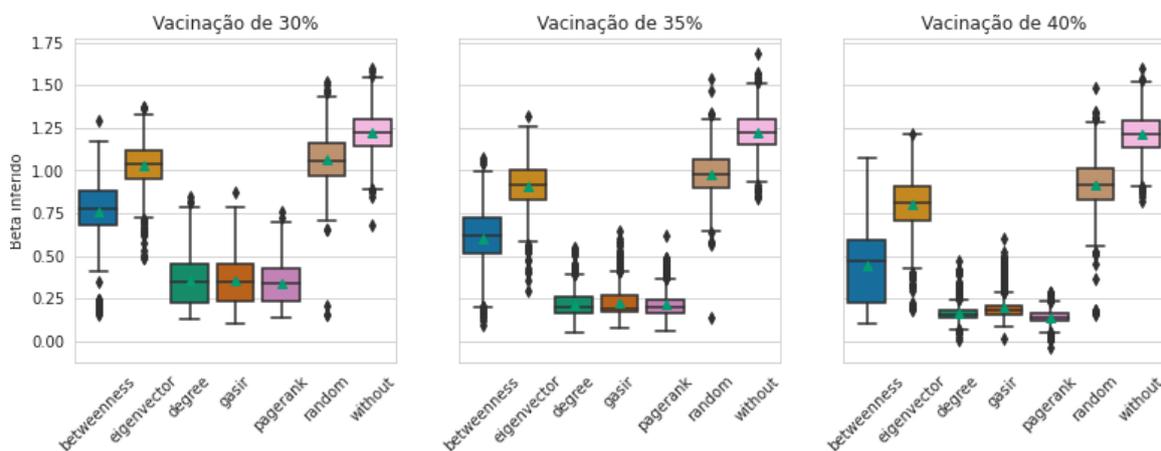
Fonte: O autor

Figura 51 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 15% e 25% na rede *innovation*.



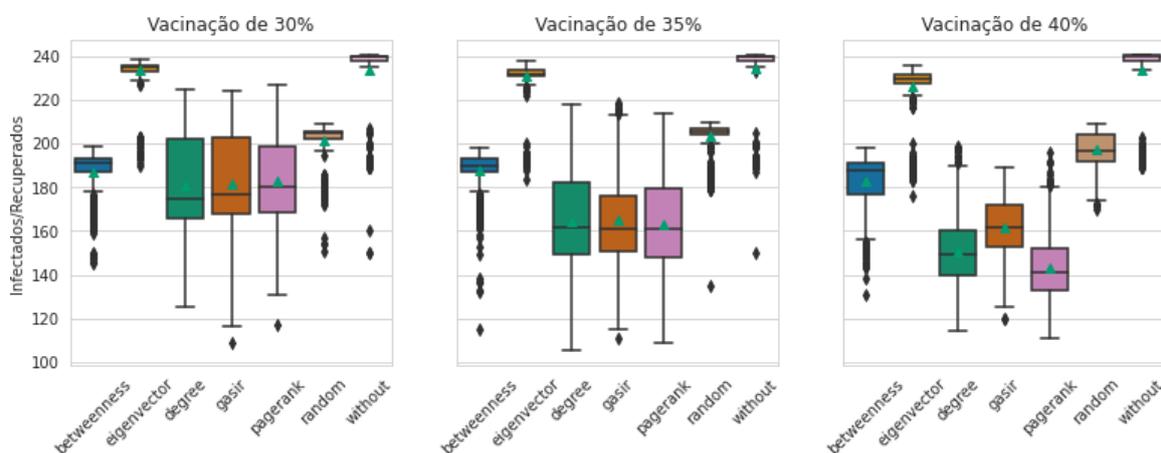
Fonte: O autor

Figura 52 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 15% e 25% na rede *innovation*.



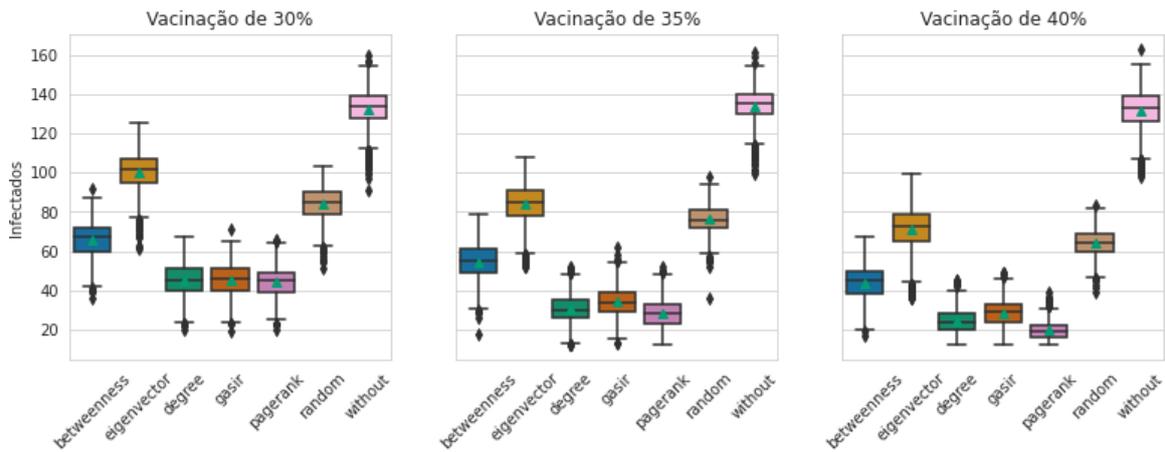
Fonte: O autor

Figura 53 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 30% e 40% na rede *innovation*.



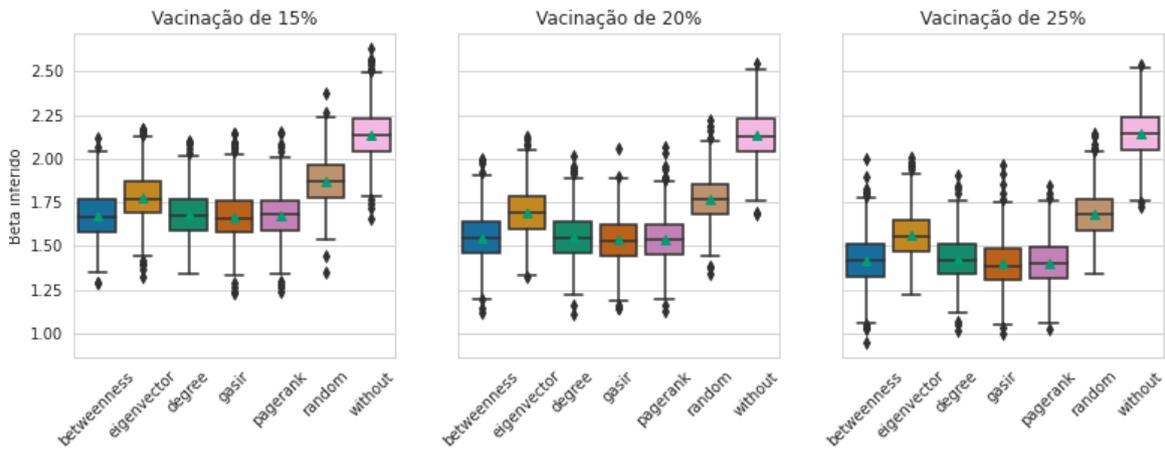
Fonte: O autor

Figura 54 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 30% e 40% na rede *innovation*.



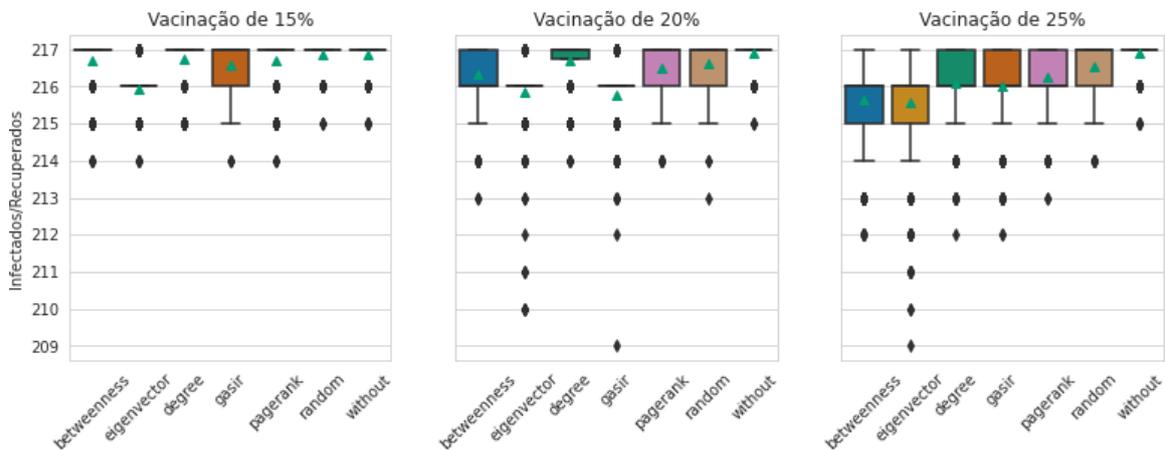
Fonte: O autor

Figura 55 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 30% e 40% na rede *innovation*.



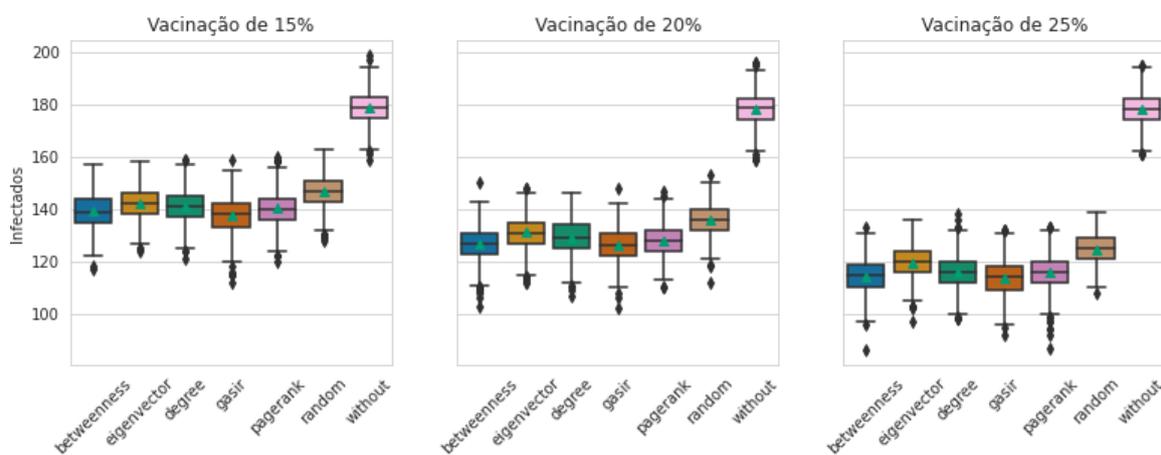
Fonte: O autor

Figura 56 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 15% e 25% na rede *oz*.



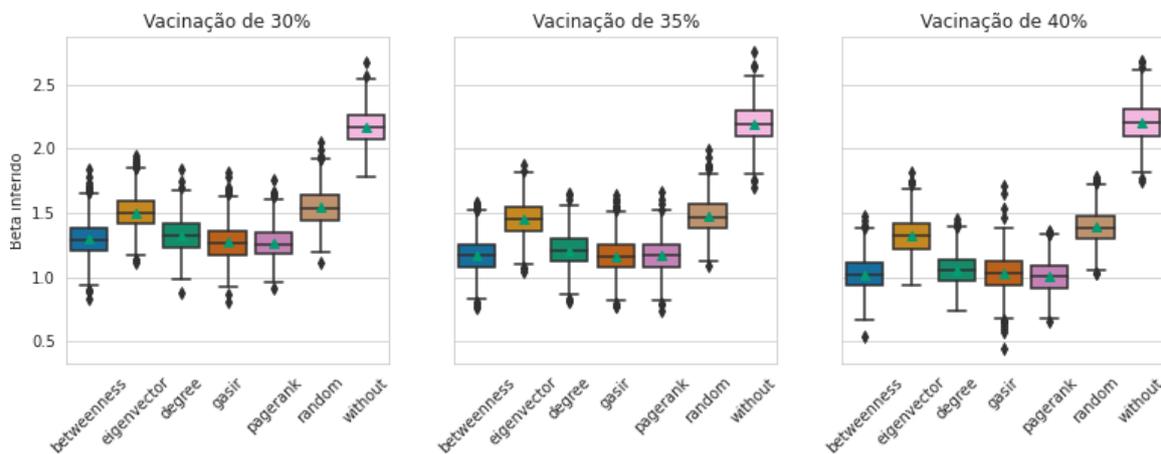
Fonte: O autor

Figura 57 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 15% e 25% na rede *oz*.



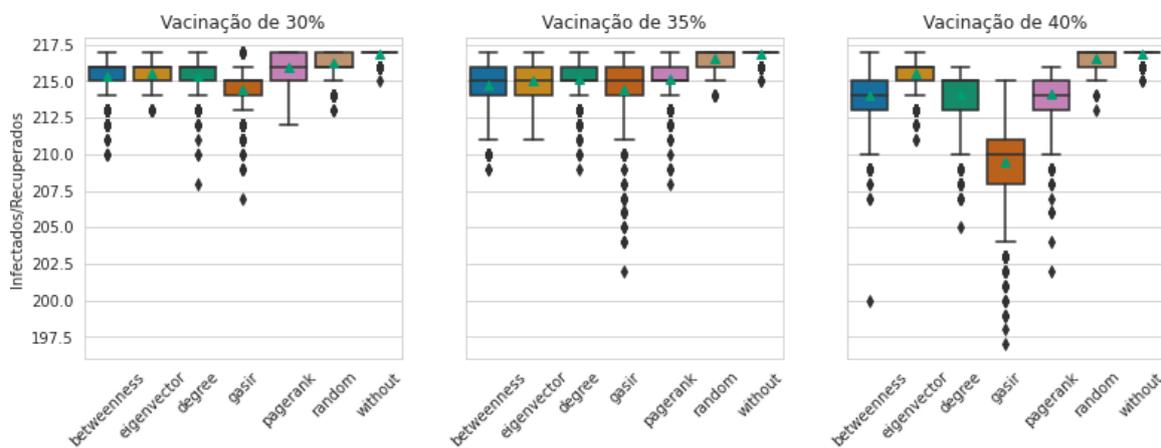
Fonte: O autor

Figura 58 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 15% e 25% na rede oz.



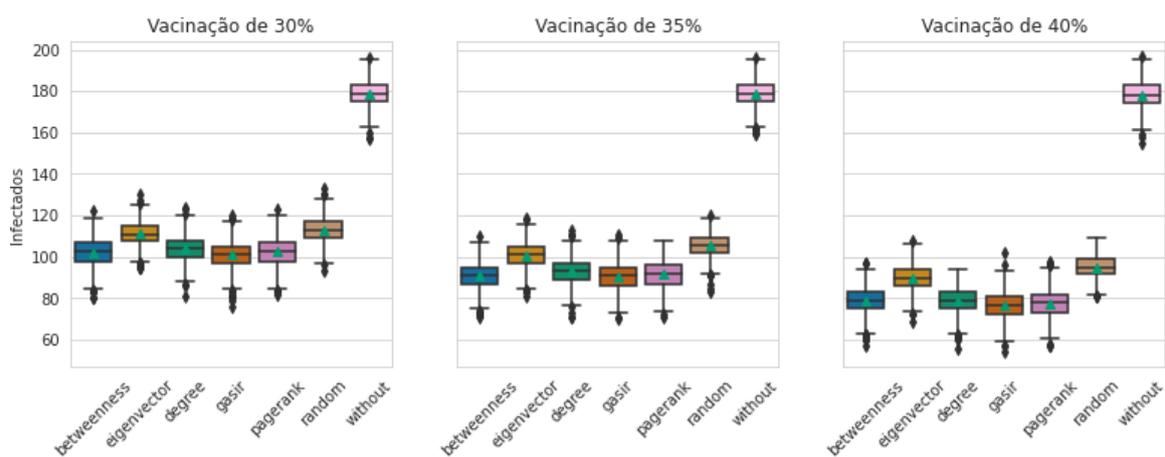
Fonte: O autor

Figura 59 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 30% e 40% na rede oz.



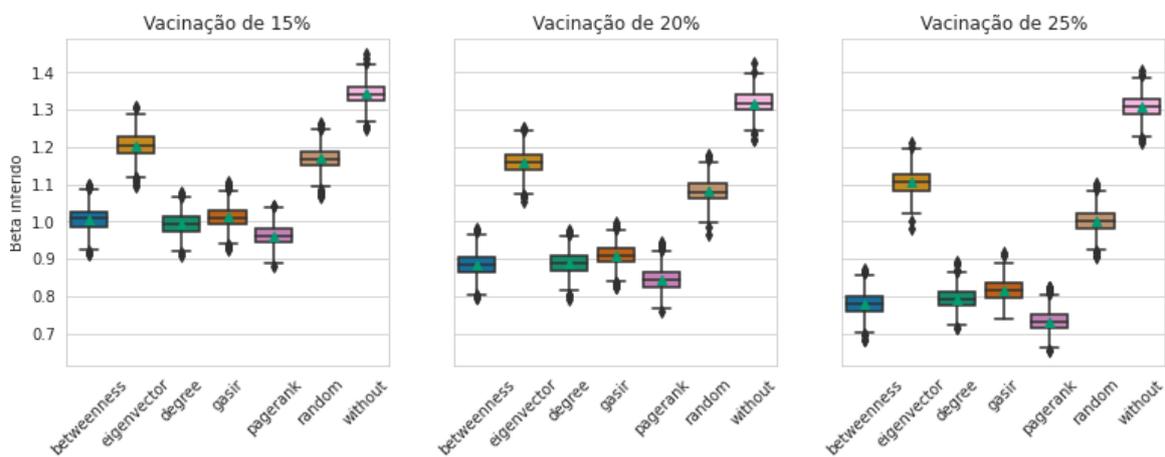
Fonte: O autor

Figura 60 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 30% e 40% na rede oz.



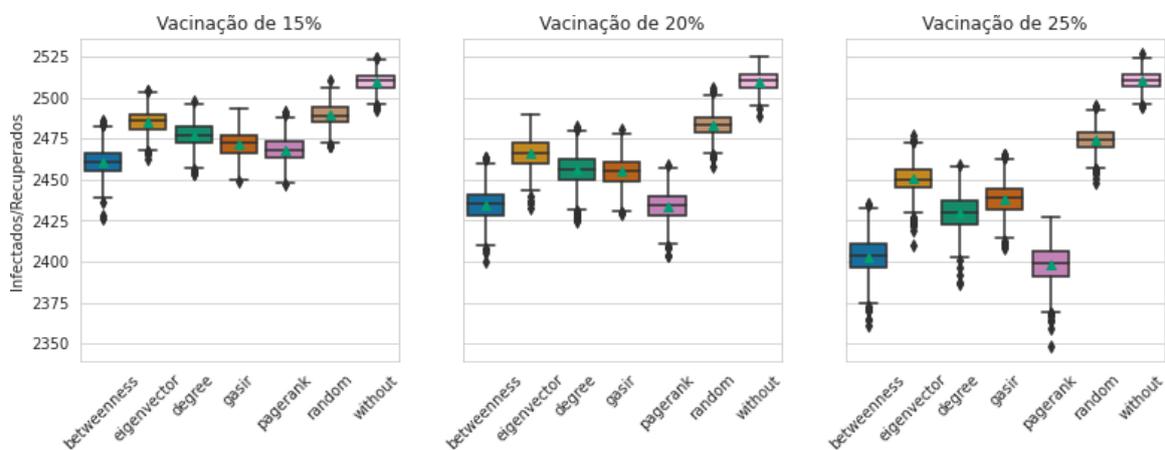
Fonte: O autor

Figura 61 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 30% e 40% na rede *oz*.



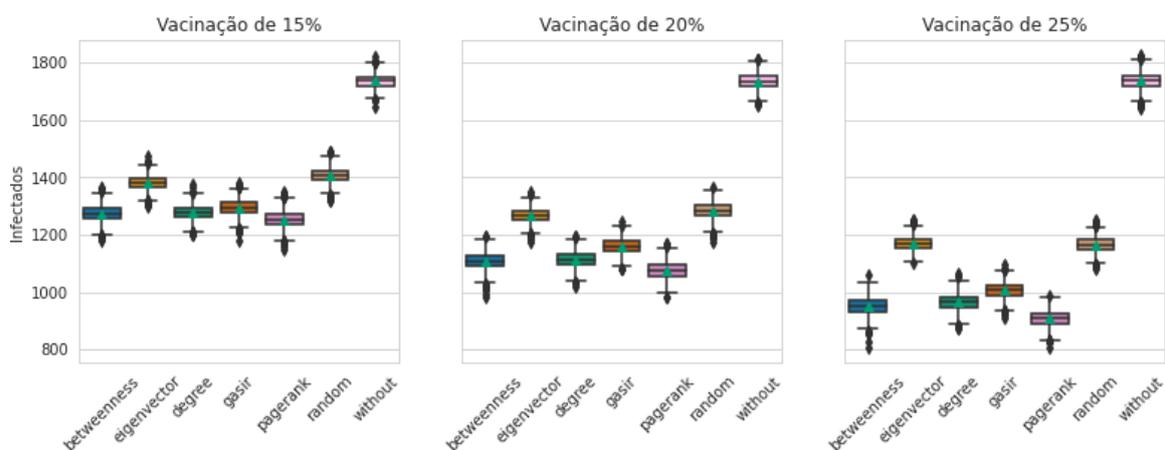
Fonte: O autor

Figura 62 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 15% e 25% na rede *health*.



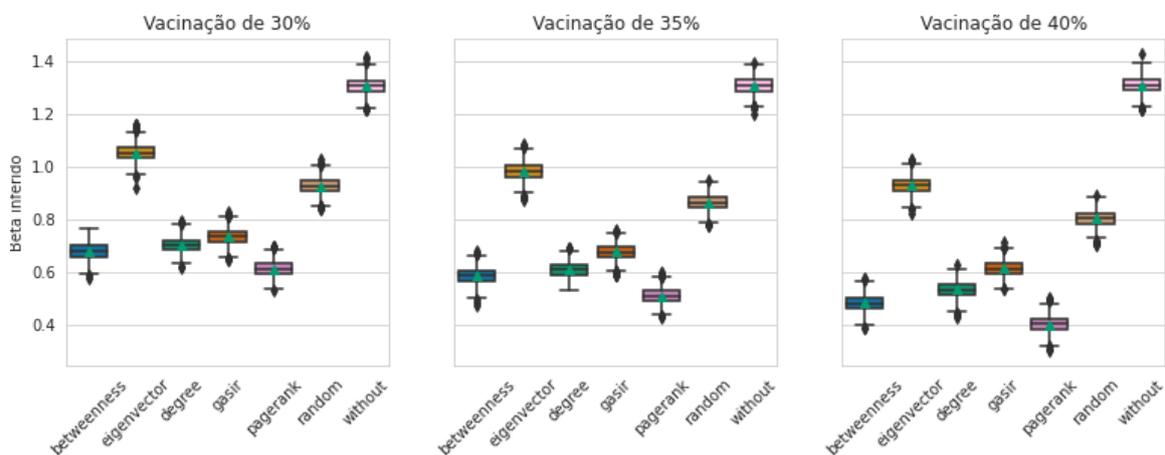
Fonte: O autor

Figura 63 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 15% e 25% na rede *health*.



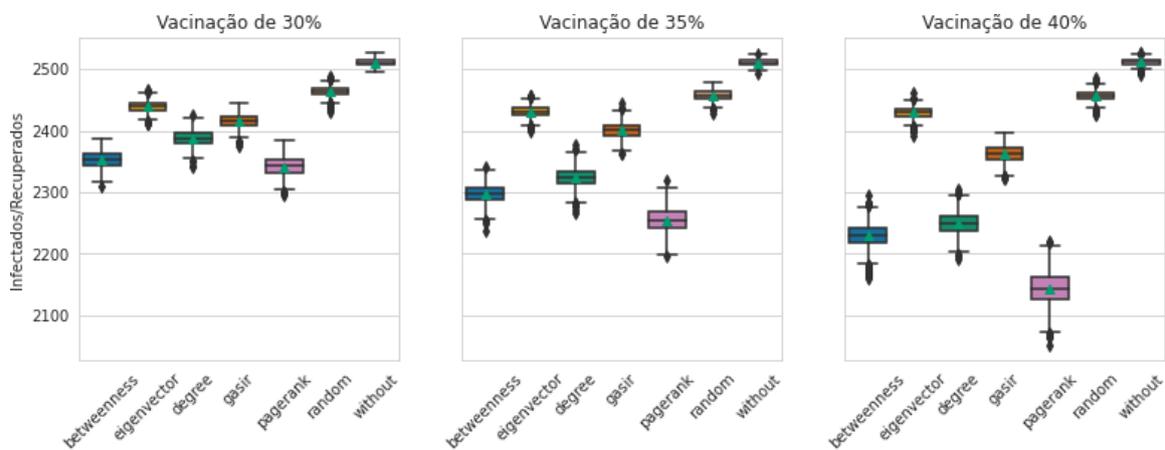
Fonte: O autor

Figura 64 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 15% e 25% na rede *health*.



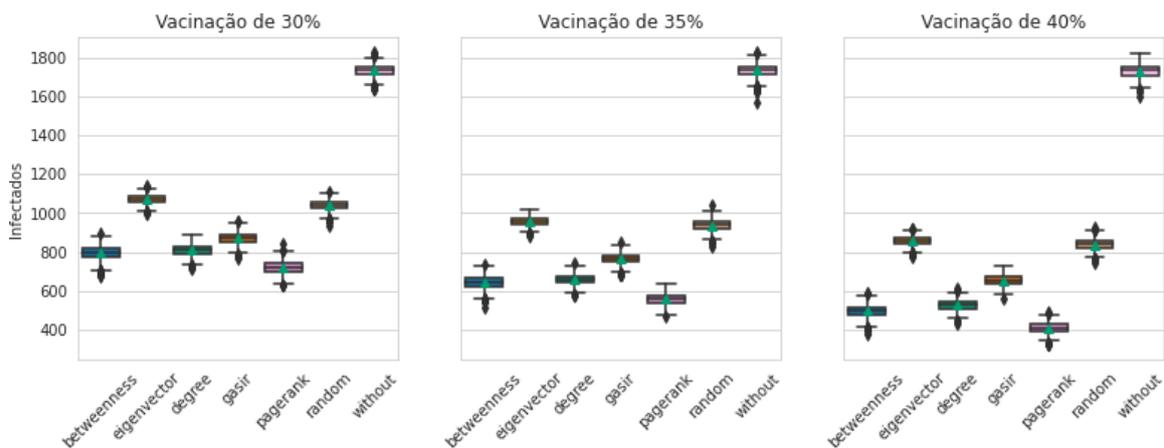
Fonte: O autor

Figura 65 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 30% e 40% na rede *health*.



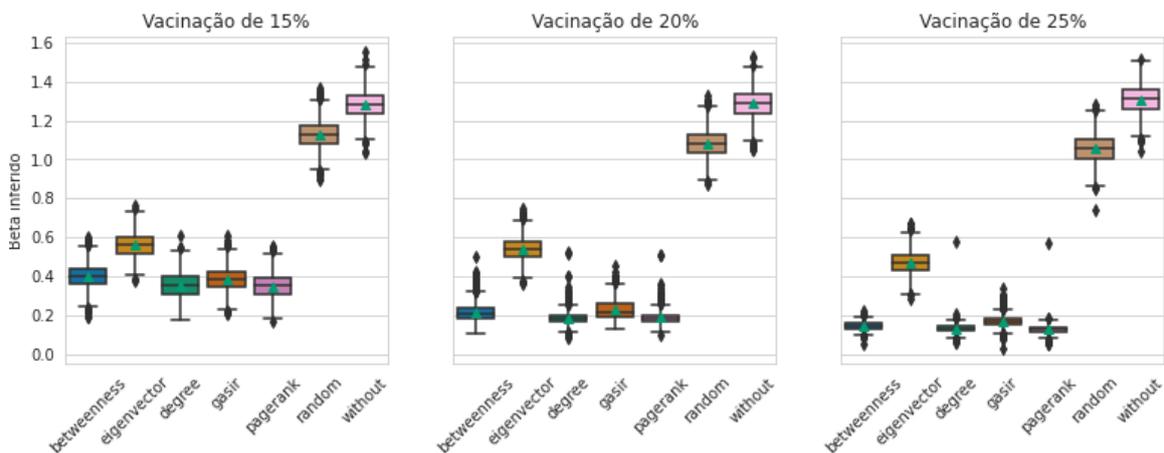
Fonte: O autor

Figura 66 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 30% e 40% na rede *health*.



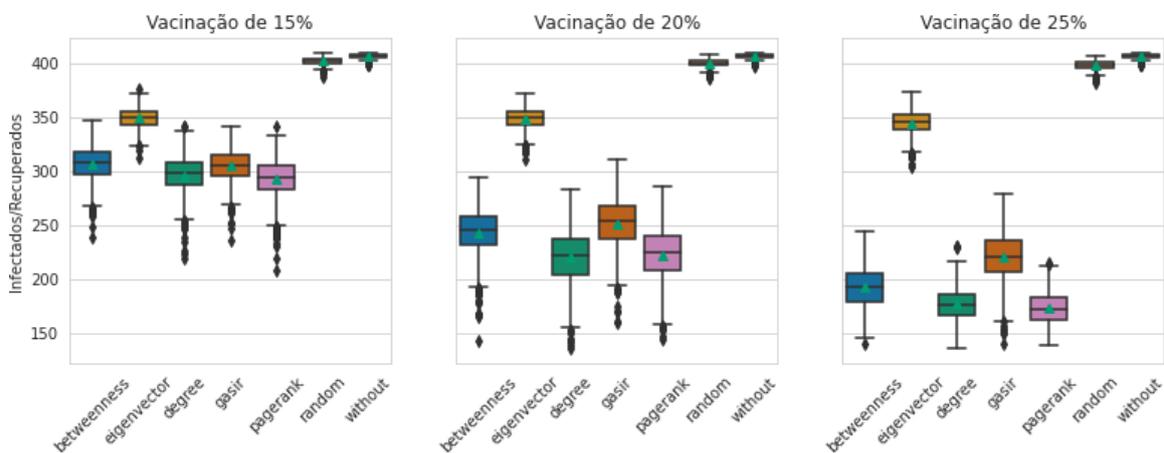
Fonte: O autor

Figura 67 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 30% e 40% na rede *health*.



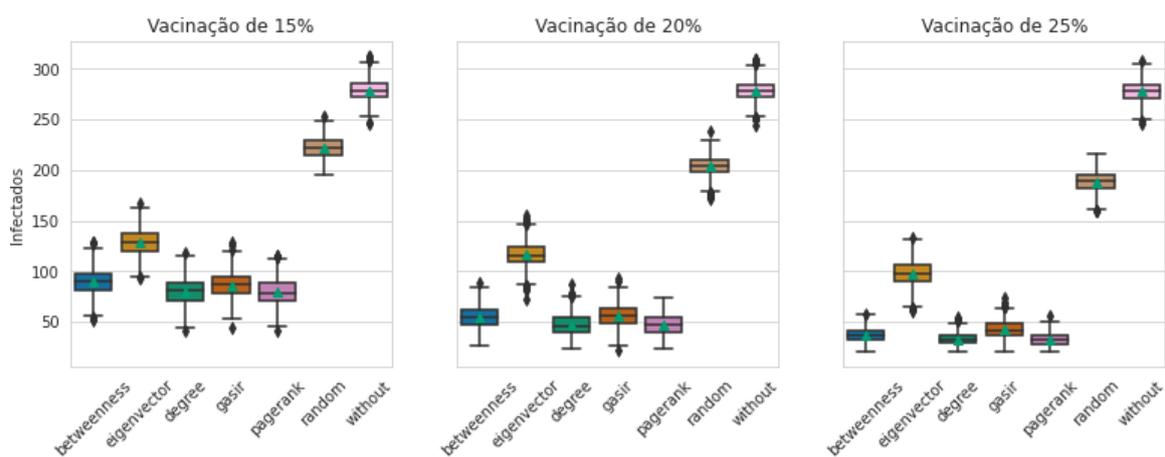
Fonte: O autor

Figura 68 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 15% e 25% na rede *ba400-m3*.



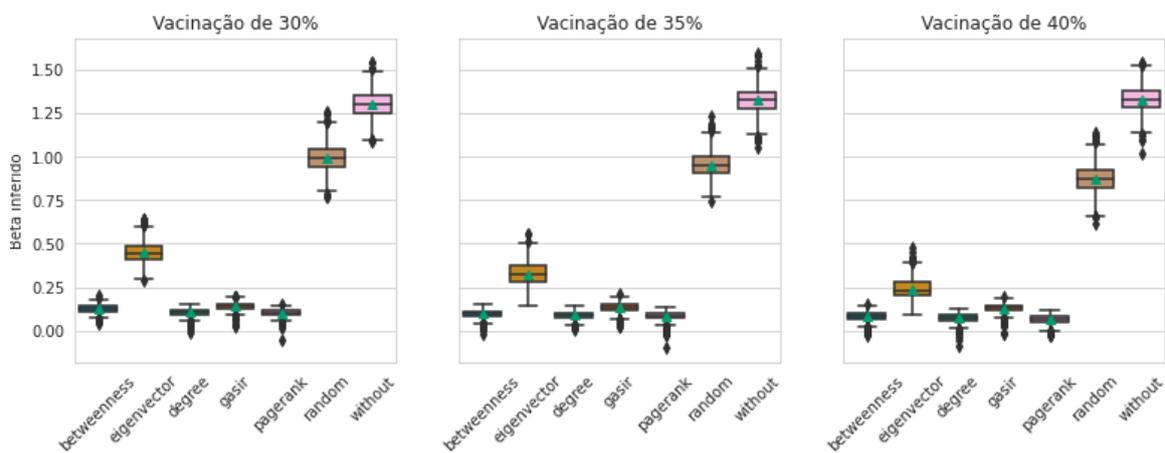
Fonte: O autor

Figura 69 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 15% e 25% na rede *ba400-m3*.



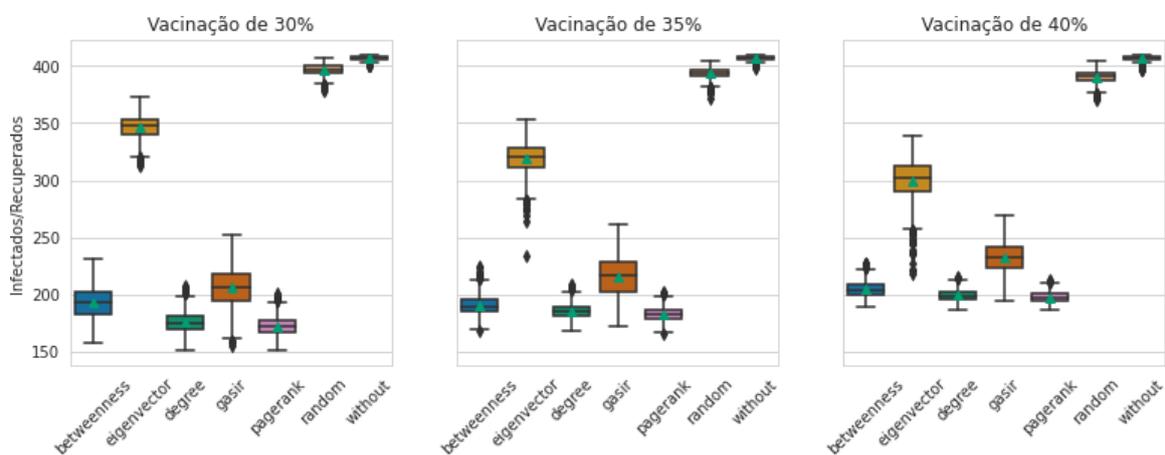
Fonte: O autor

Figura 70 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 15% e 25% na rede *ba400-m3*.



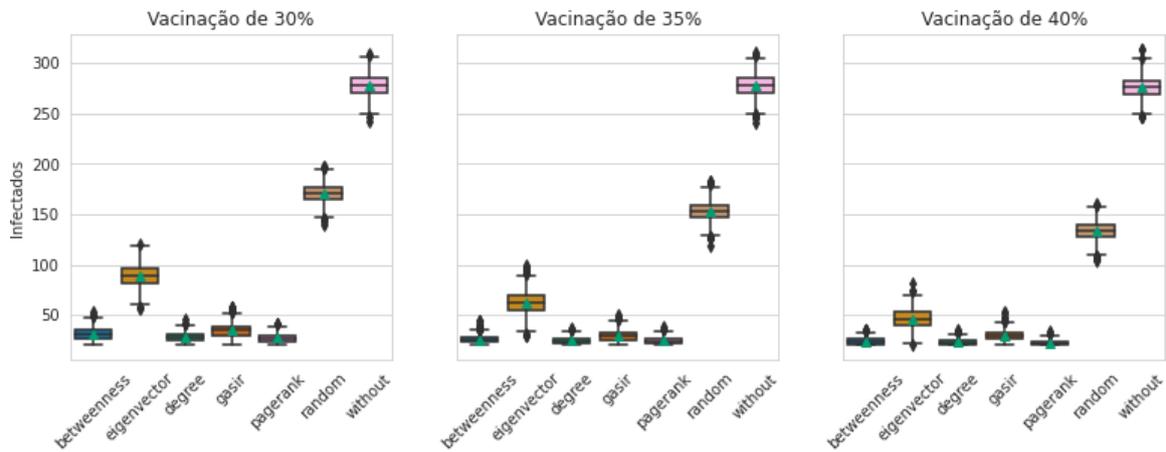
Fonte: O autor

Figura 71 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 30% e 40% na rede *ba400-m3*.



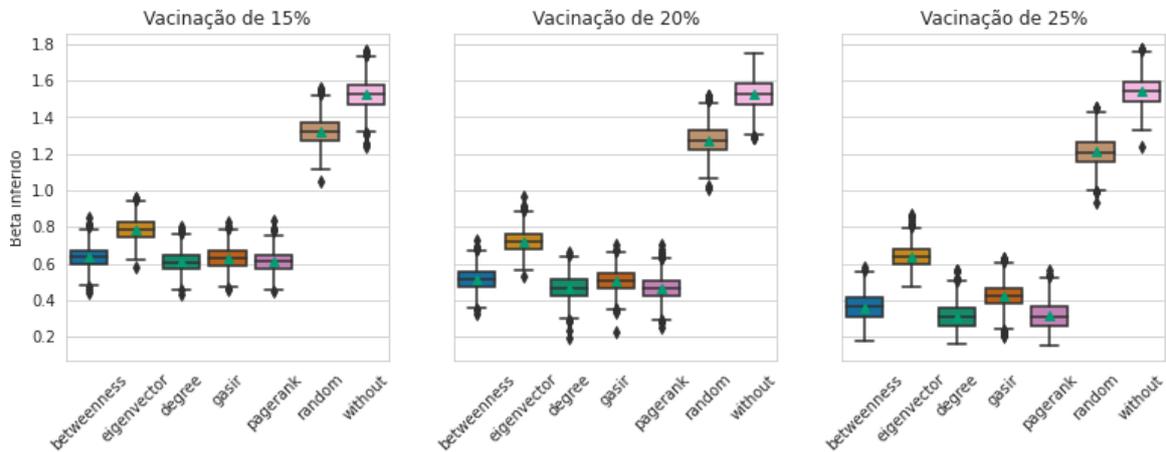
Fonte: O autor

Figura 72 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 30% e 40% na rede *ba400-m3*.



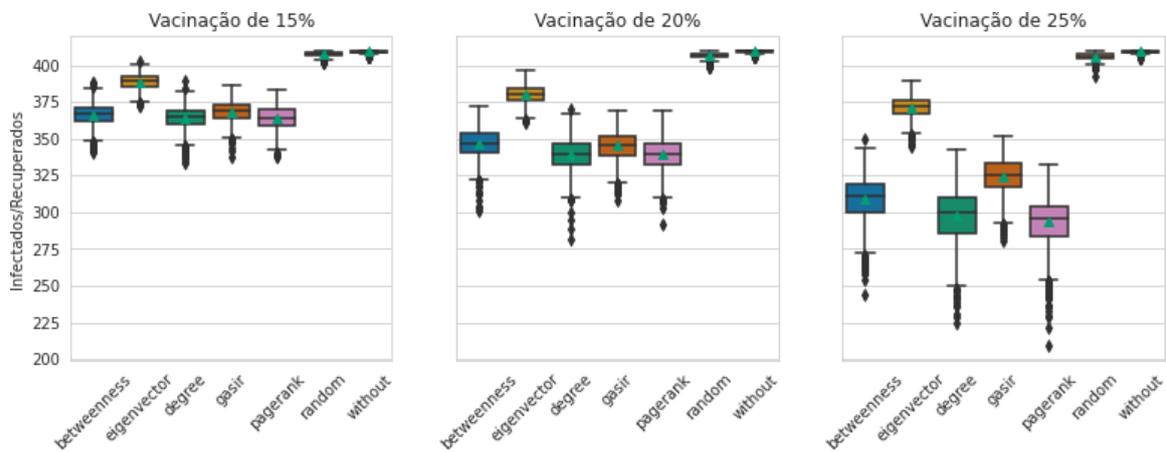
Fonte: O autor

Figura 73 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 30% e 40% na rede *ba400-m3*.



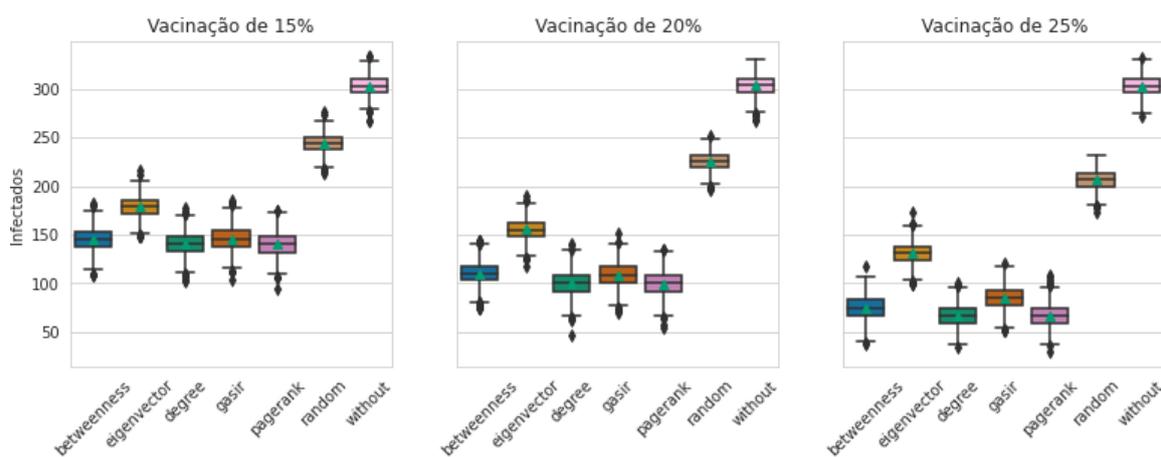
Fonte: O autor

Figura 74 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 15% e 25% na rede *ba400-m4*.



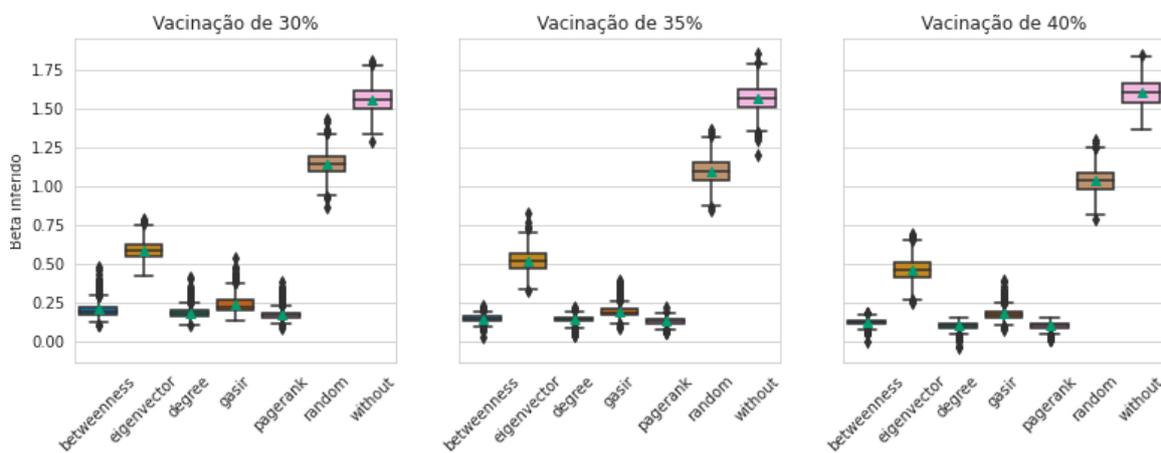
Fonte: O autor

Figura 75 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 15% e 25% na rede *ba400-m4*.



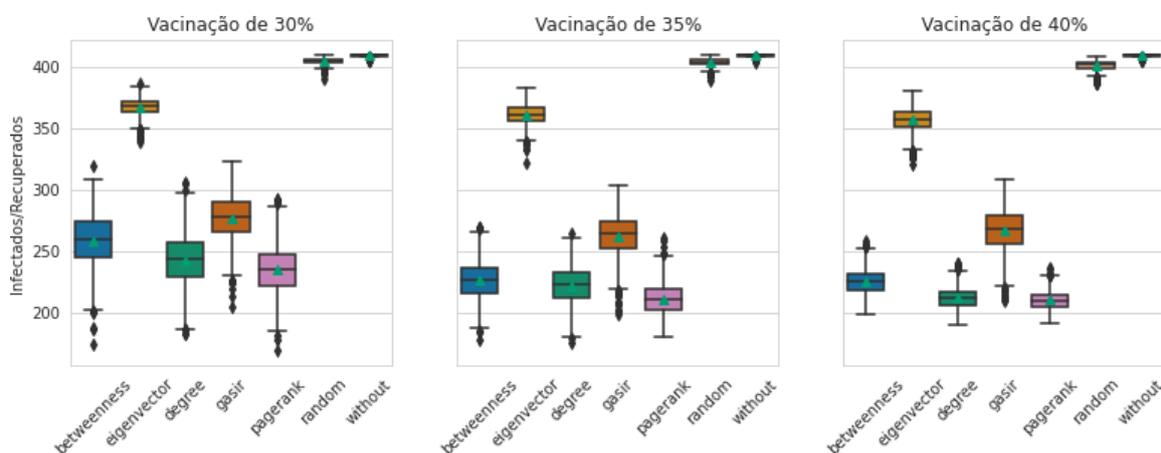
Fonte: O autor

Figura 76 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 15% e 25% na rede *ba400-m4*.



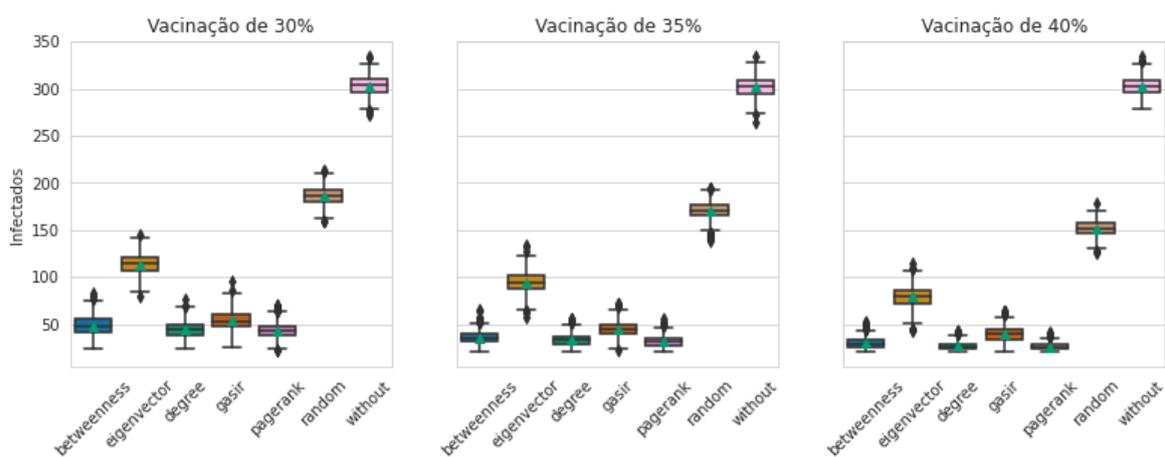
Fonte: O autor

Figura 77 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 30% e 40% na rede *ba400-m4*.



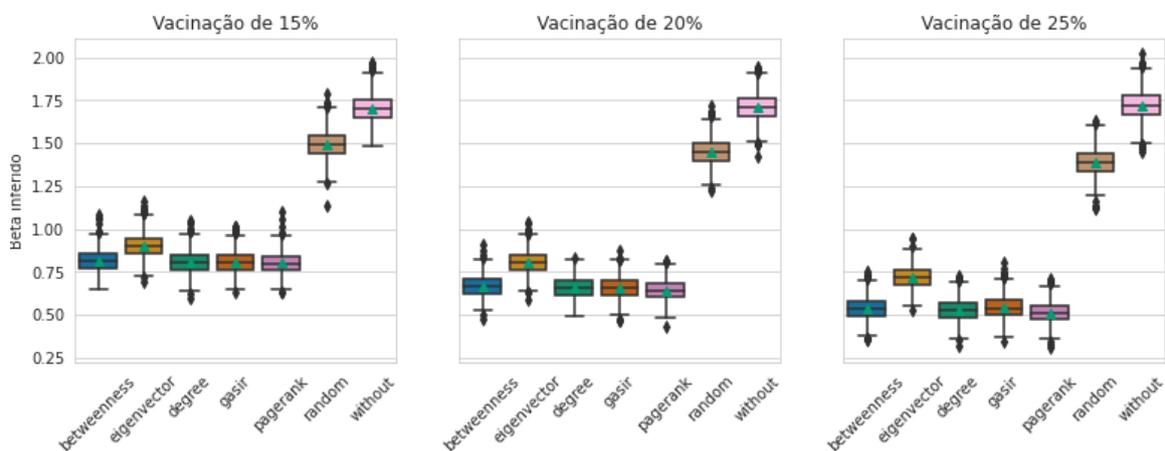
Fonte: O autor

Figura 78 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 30% e 40% na rede *ba400-m4*.



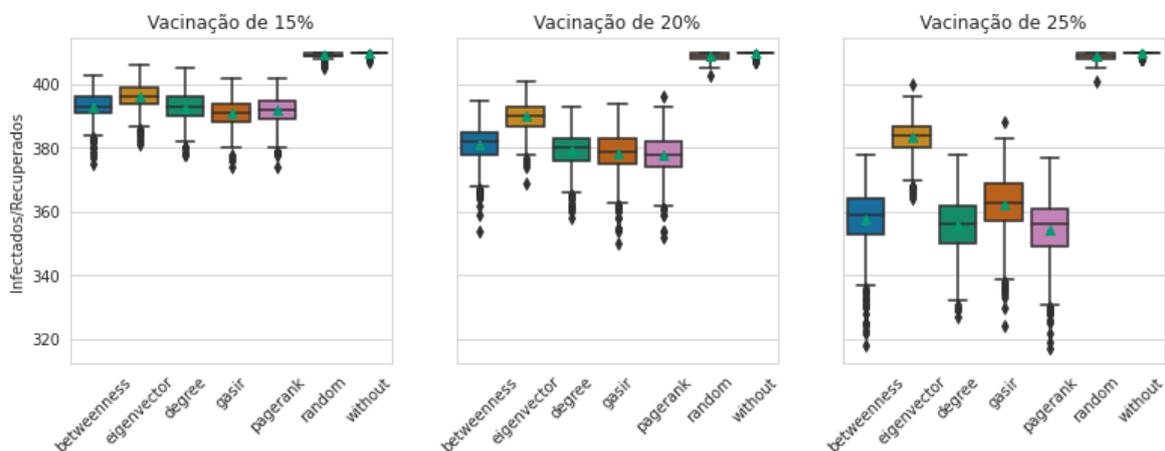
Fonte: O autor

Figura 79 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 30% e 40% na rede *ba400-m4*.



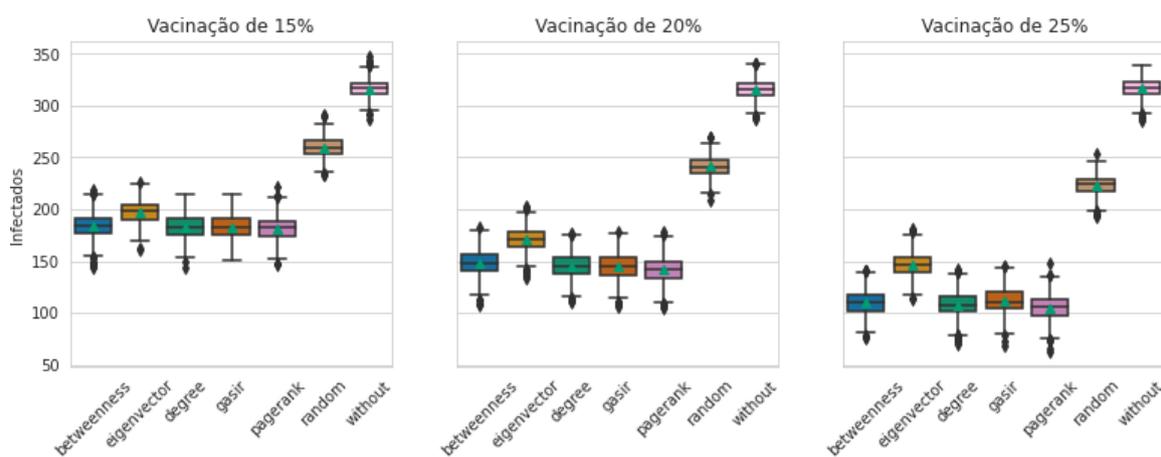
Fonte: O autor

Figura 80 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 15% e 25% na rede *ba400-m5*.



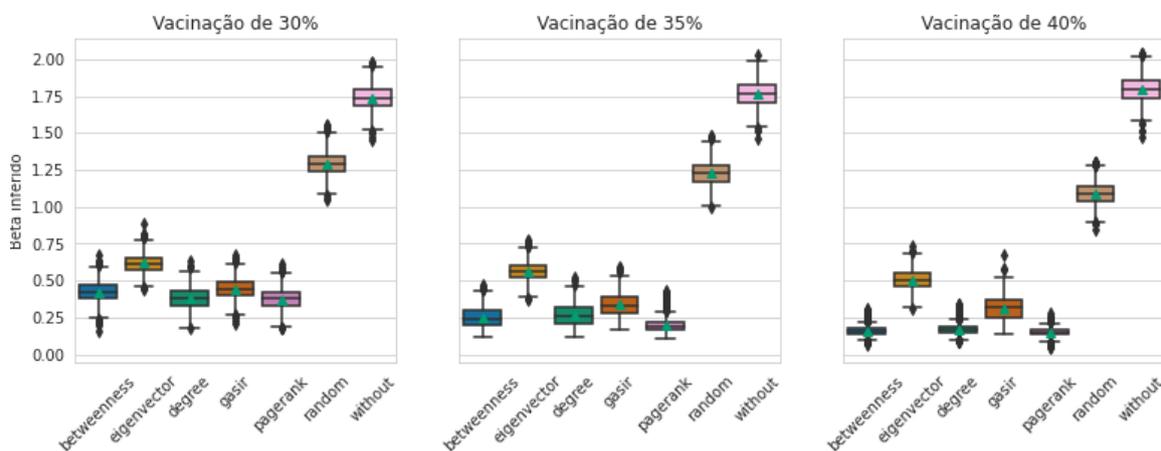
Fonte: O autor

Figura 81 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 15% e 25% na rede *ba400-m5*.



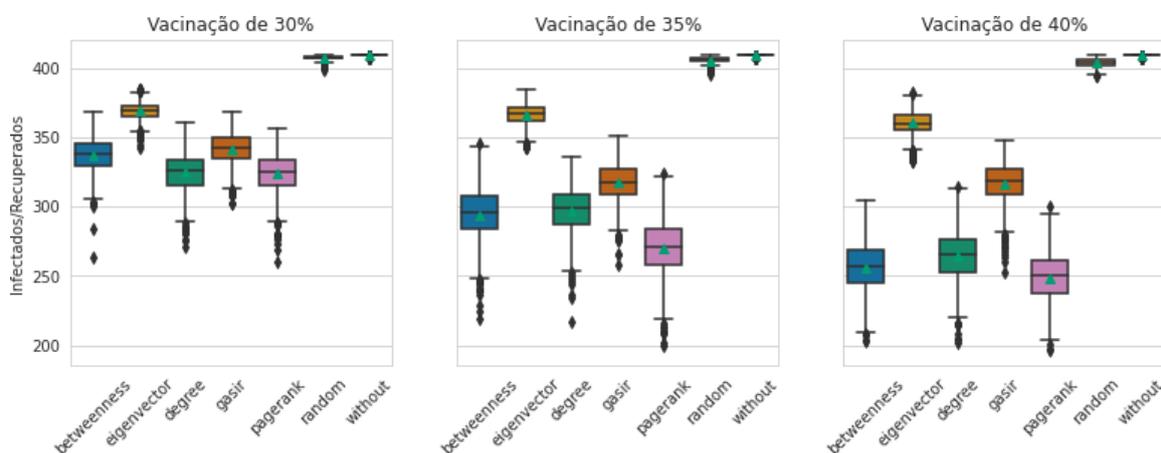
Fonte: O autor

Figura 82 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 15% e 25% na rede *ba400-m5*.



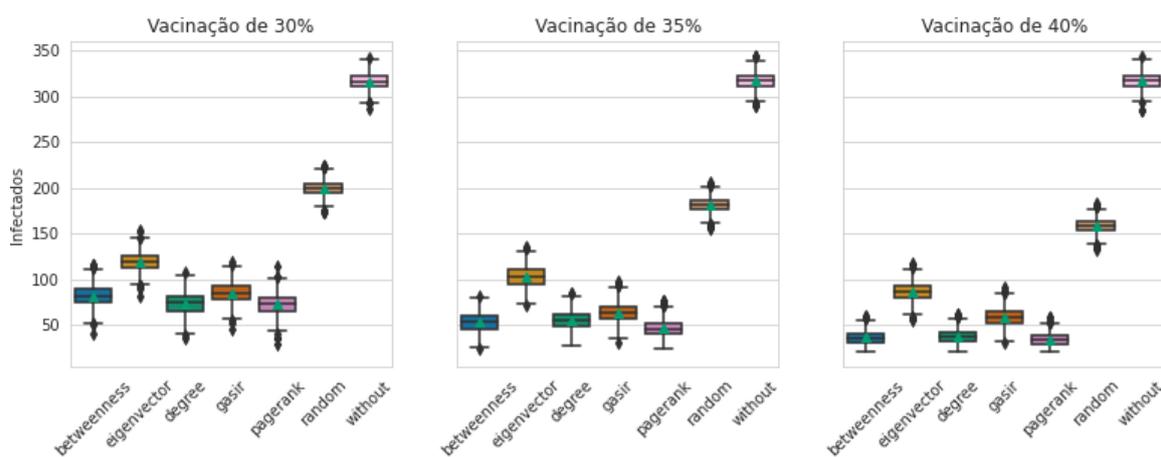
Fonte: O autor

Figura 83 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 30% e 40% na rede *ba400-m5*.



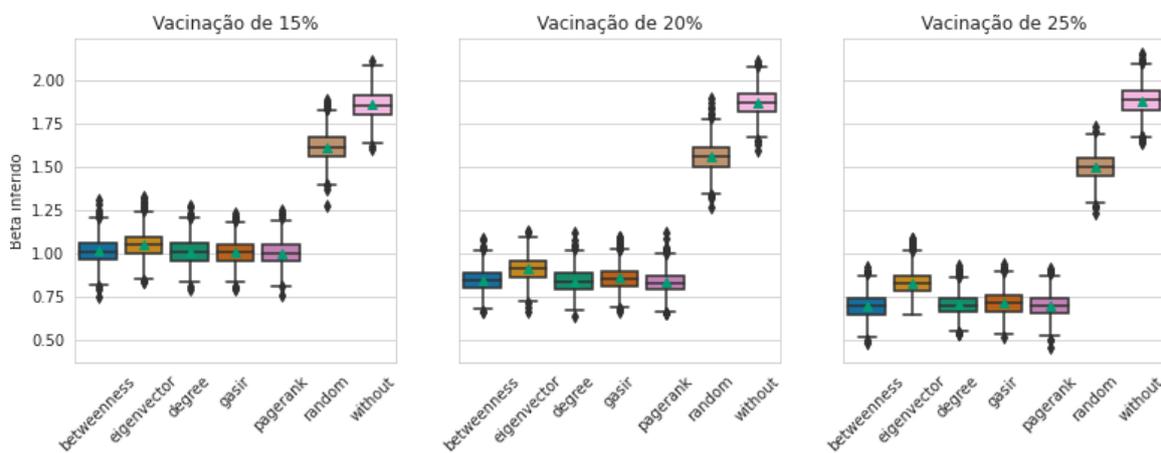
Fonte: O autor

Figura 84 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 30% e 40% na rede *ba400-m5*.



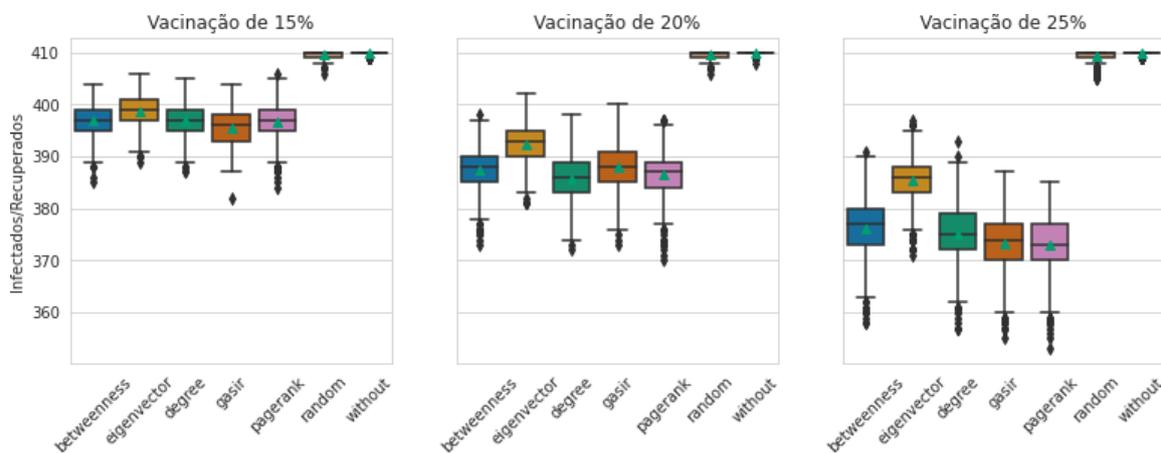
Fonte: O autor

Figura 85 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 30% e 40% na rede *ba400-m5*.



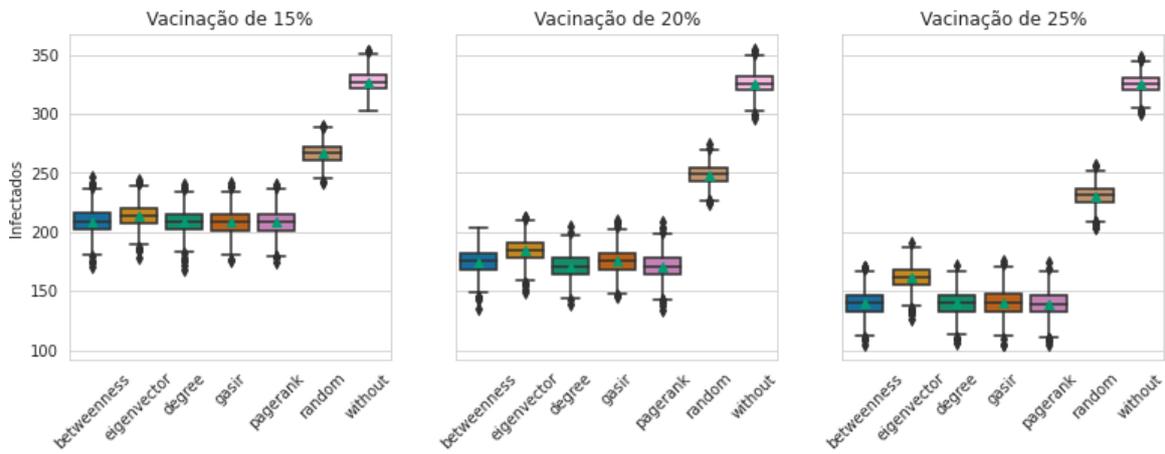
Fonte: O autor

Figura 86 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 15% e 25% na rede *ba400-m6*.



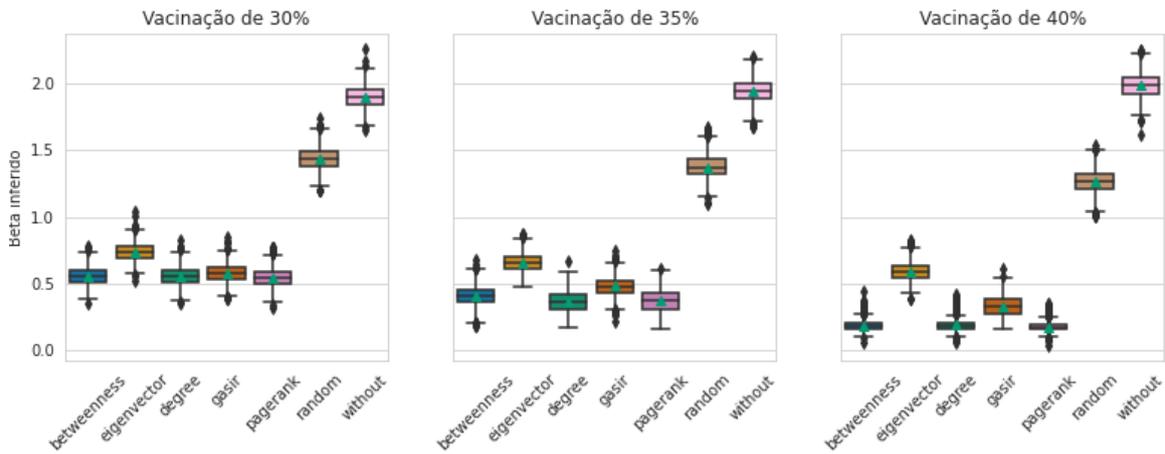
Fonte: O autor

Figura 87 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 15% e 25% na rede *ba400-m6*.



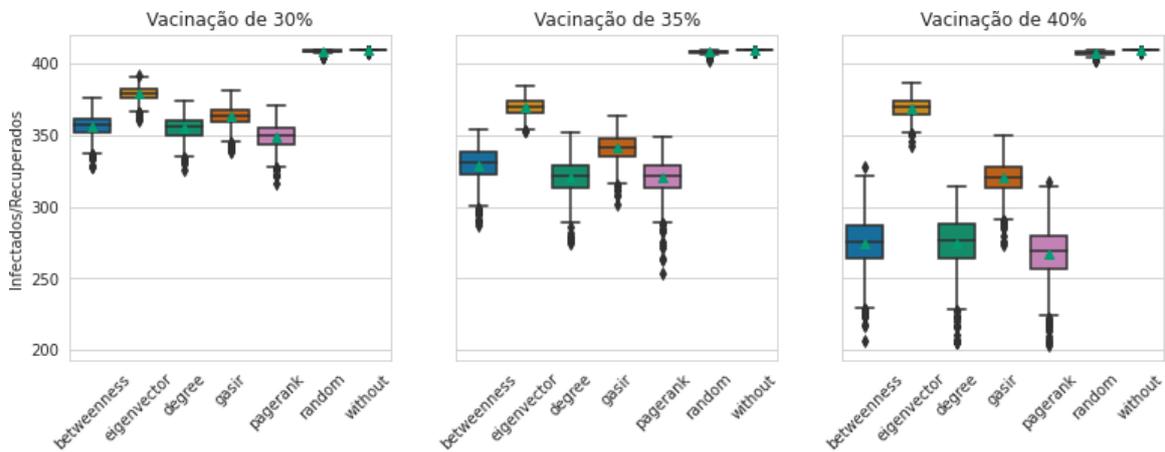
Fonte: O autor

Figura 88 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 15% e 25% na rede *ba400-m6*.



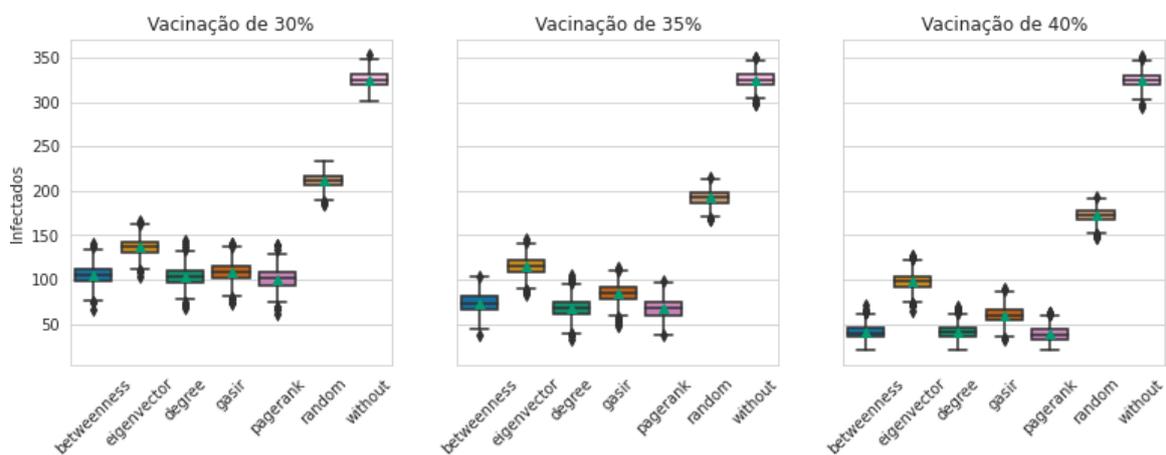
Fonte: O autor

Figura 89 – Boxplot com os resultados obtidos do beta inferido com vacinação entre 30% e 40% na rede *ba400-m6*.



Fonte: O autor

Figura 90 – Boxplot com os resultados obtidos do número de indivíduos afetados com vacinação entre 30% e 40% na rede *ba400-m6*.



Fonte: O autor

Figura 91 – Boxplot com os resultados obtidos do pico de infectados com vacinação entre 30% e 40% na rede *ba400-m6*.