

# Ajuste do modelo SEIR para simulação do espalhamento do Corona Vírus

Carolina Xavier - Tulio

September 17, 2020

## 1 Introdução

Modelos epidemiológicos são amplamente utilizados na literatura para a análise e simulação de comportamentos de epidemias, entre eles, temos o modelo SEIR (Suscetível/Exposto/Infetado/Removido). Os modelos epidemiológicos clássicos compartimentalizam a população para descrever a dinâmica de infecção de uma epidemia, estes modelos são baseados em equações diferenciais e não levam em consideração as relações e contato entre indivíduos, considerando a mesma possibilidade de qualquer indivíduo ser contaminado, é interessante então, utilizar uma variação do modelo original, que tem sua dinâmica baseada em redes complexas [1]. No modelo SEIR em redes a infecção se dá através do contato de indivíduos, levando em consideração a probabilidade de infecção e/ou imunidade de um indivíduo e probabilidade de recuperação, além do período de incubação do vírus.

Cada doença tem diferentes dinâmicas e portanto, elas possuem diferentes taxas de probabilidade de infecção e recuperação, sendo necessário ajustar esses valores para cada nova doença contagiosa. Ajustados esses parâmetros é possível simular os efeitos da epidemia para a doença específica em diferentes cenários.

Dentro das diversas heurísticas existentes para a resolução de problemas de otimização, os algoritmos genéticos são um dos destaques, devido a sua facilidade de implementação e resultados finais satisfatórios. Em face disto, este trabalho propõe uma metodologia ajuste dos parâmetros do modelo SIR através de algoritmos genéticos a fim de simular o espalhamento da epidemia do corona Vírus e permitir o estudo de outras epidemias com comportamento similar.

Modelos epidemiológicos descrevem de forma simplificada a dinâmica de transmissão de epidemias. Tais modelos podem ser divididos em duas categorias, estocásticos e determinísticos. Modelos estocásticos estimam a probabilidade de infecções, permitindo a variação de forma aleatória de suas entradas no decorrer de sua execução [2]. Já em modelos determinísticos, a população (geralmente grande) é subdividida em grupos ou compartimentos onde cada grupo representa um estágio da epidemia. A dinâmica de tais grupos é descrita por equações diferenciais.

### Modelo Epidemiológico SIR

Desenvolvido por Kermack e McKendrick [3], este modelo considera uma população fixa e a divide em 3 grupos:

- Suscetíveis: Indivíduos que podem ser infectados por outros indivíduos.
- Infectados: Indivíduos infectados que podem infectar outros indivíduos suscetíveis.
- Recuperados ou Removidos: Indivíduos que já foram infectados mas se recuperaram (ou vieram ao óbito). São resistentes a doença e não podem infectar e nem mais serem infectados.

Este modelo segue um fluxo unidirecional entre os 3 grupos. Cada grupo representa um estágio onde um indivíduo ou grupo de indivíduos se encontram em relação à epidemia. Vale ressaltar que o indivíduo só pode pertencer a apenas um grupo por vez.

O modelo SEIR em sua forma original não leva em conta a interação entre os indivíduos, apenas descreve o tamanho de cada grupo em um instante de tempo  $t$  e estabelece como são as funções de transições entre os grupos.

Seja uma população fixa  $N = S(t) + I(t) + R(t)$ , Kermack e McKendrick [3] descrevem a densidade do grupo Suscetível, Infectado, Recuperado, respectivamente, em um instante  $t$  através das seguintes equações diferenciais:

$$dS/dt = -\beta SI \quad dI/dt = \beta SI - \gamma I \quad dR/dt = \gamma I \quad (1)$$

sendo  $\beta$  a velocidade de contato ou infecção da doença e  $\gamma$  a taxa média de recuperação/óbito.

O modelo SEIR engloba ainda o tempo de incubação da doença.

### Modelo SEIR sobre redes

Além do modelo de equações diferenciais, o modelo SEIR também pode ser implementado através de uma rede de contatos [4]. Desta forma, cada nó na rede representa um indivíduo na população e cada aresta entre os indivíduos representa um contato interpessoal, através do qual a doença pode ser transmitida.

No modelo SEIR [5], durante o curso de uma epidemia, um nó pode mudar seu status de Suscetível (S) para Exposto (E) para Infetado (I), depois para Removido (R).

O SEIR assume que se, durante uma iteração genérica, um nó suscetível entrar em contato com um infectado, ele será infectado após um período  $\alpha$  de exposição com probabilidade  $\beta$ , do que poderá mudar para removido com probabilidade  $\gamma$  (a única transição permitida é S  $\rightarrow$  E  $\rightarrow$  I  $\rightarrow$  R).

Para novas doenças esses valores precisam ser ajustados, para isso utilizaremos dados públicos do quantitativo de infectados, recuperados e óbitos por covid19 para ajustar os valores de  $\alpha, \beta$  e  $\gamma$ . *Para esses ajustes serão utilizadas redes complexas sintéticas.*

Algoritmos Genéticos (AG) são uma subclasse dos algoritmos evolutivos. Os AGs utilizam técnicas inspiradas na teoria da evolução, como seleção natural, hereditariedade, recombinação e mutação [6]. Constituem uma heurística poderosa devido ao seu poder de abstração e baixo consumo de recursos computacionais, devido a este fato, são amplamente utilizados em problemas de busca e otimização.

Baseado no processo de seleção natural proposto por Charles Darwin e Alfred Wallace em [7], tem-se o pressuposto que em uma população constituída de indivíduos com características diversas, os indivíduos mais adaptados ao meio possuem mais chances de perpetuar suas características através do processo de reprodução. Em face disto, para tal simulação, algoritmos genéticos são modelados da seguinte forma:

- Gene: Características que compõe um indivíduo, geralmente assumem valores binários ou valores inteiros positivos.
- Cromossomo ou Indivíduo: Conjunto de genes ou características, representa uma solução no problema.
- População: Conjunto de cromossomos ou soluções candidatas para o problema proposto.

O algoritmo genético submete a população a 4 etapas a cada geração. São elas:

- Avaliação: Através de uma função avaliadora, cada indivíduo (ou solução) é avaliado e uma pontuação (fitness) é atribuída a ele.
- Seleção de Pais: Nesta etapa um conjunto de indivíduos é selecionado para realizarem o cruzamento de seus genes. Indivíduos com uma pontuação maior têm maiores chances de serem selecionados. Esta etapa pode ser realizada de diversas formas, como roletas, torneios ou até mesmo de forma aleatória.
- Cruzamento: De posse do conjunto de pais selecionados, dois pais são sorteados e é realizado o cruzamento de seus genes. Um dos tipos de cruzamentos utilizados é o cruzamento de n-pontos, podendo ser escolhidos o número de pontos de cruzamento de acordo com o problema a ser resolvido. Este processo é repetido até que se obtenha uma nova população do tamanho da população original.

Esta etapa é muito importante, pois é responsável por combinar soluções, podendo produzir soluções ainda melhores, garantindo a intensificação da busca.

- Mutação: Com a nova população gerada, a etapa de mutação tem por finalidade a exploração por novas soluções. Este processo evita que o algoritmo fique preso em soluções ótimas locais, impedindo que o mesmo fique estagnado. A mutação ocorre de forma aleatória e com baixa probabilidade de ocorrência. Seu processo é simples, o algoritmo percorre toda a sua população, e para cada gene de cada indivíduo é gerado um número aleatório, se este número atingir o limiar de mutação o valor do gene atual é alterado para um valor válido aleatório.

Estas 4 etapas são repetidas até que o número máximo de gerações definidas seja alcançado. A cada geração novos indivíduos são gerados a partir de pais da geração atual que apresentaram boas soluções para o problema, causando assim uma convergência para um grupo de indivíduos candidatos à solução ao final da execução.

## 2 Objetivos

O objetivo deste projeto é ajustar o modelo SIR para o funcionamento satisfatório da simulação do espalhamento do corona Vírus no Brasil, para que diferentes cenários possam ser testados e investigados.

Os objetivos específicos são:

- O entendimento no modelo SIR em redes e cada um de seus parâmetros;
- O entendimento no modelo SEIR em redes e cada um de seus parâmetros;
- Construção de uma função objetivo baseada no modelo SIR;
- Construção de uma função objetivo baseada no modelo SEIR;
- Construção de um algoritmo de otimização para ajuste dos parâmetros;
- Análise dos resultados do algoritmo de otimização;
- Uso do modelo para simulação da epidemia em diferentes cenários.

## 3 Metodologia

A metodologia a ser implementada neste projeto é experimental e visa seguir os passos descritos a seguir para obtenção de um conjunto de parâmetros que melhor se ajusta as curvas reais de suscetíveis, infectados e recuperados/óbitos por covid19.

- Determinação da função objetivo como diferença entre os valores obtidos pelo modelo SIR e os valores reais da epidemia por covid19 no Brasil;
- Implementação de um AG para otimização da função objetivo criada;
- Experimento fatorial completo para escolha do conjunto de parâmetros do AG que melhor otimiza a função objetivo;
- Comparação com outros trabalhos disponíveis na literatura;
- Publicação de resultados.

## 4 Plano de trabalho

O bolsista irá estudar a base teórica para o desenvolvimento do trabalho para então desenvolver os passos da metodologia apresentada anteriormente. O cronograma de trabalho do aluno será da forma:

- Estudo de modelos epidemiológicos- Meses 1 e 2
- Entendimento dos parâmetros do modelo SIR e SEIR - Meses 2 e 3
- Estudo de algoritmos genéticos e seus parâmetros - Meses 2, 3 e 4
- Implementação do AG - Meses 3, 4 e 5
- Estudo dos parâmetros do AG (experimento fatorial completo) - Meses 5, 6 e 7
- Execução de testes com o melhor conjunto de parâmetros encontrado - Meses 6 e 7
- Estudo estatístico do resultado dos testes - Meses 7 e 8
- Redação dos resultados - Meses 8 e 9
- Comparação dos resultados com resultados de modelos utilizando EDO's e modelos de outros autores que utilizam redes complexas - Meses 9, 10 e 11

- Redação de artigos científicos e relatórios das atividades desenvolvidas - Meses 11 e 12.

Serão utilizados diversos mecanismos para acompanhamento dos bolsistas:

- Formais: O mecanismo formal básico de acompanhamento do bolsista é a cobrança da assinatura do ponto do mesmo e o acompanhamento das notas do bolsista, já que o aluno deve entender a importância de se manter a excelência acadêmica.

- Reuniões: O bolsista será acompanhado por reuniões semanais, quando as atividades realizadas serão discutidas e será também feito o planejamento das tarefas a serem realizadas.

- Relatórios: O bolsista apresentará relatório de progresso trimestral onde serão sumarizadas as atividades realizadas. Este relatório será contrastado com o cronograma de atividades.

- Artigos: O projeto prevê a submissão de artigos para conferências especializadas quando teremos a oportunidade de obter avaliações externas sobre a relevância e qualidade do trabalho que está sendo realizado.

O desenvolvimento do bolsista será avaliado pelos seguintes critérios:

- Assiduidade e Pontualidade: A disciplina é um aspecto fundamental do sucesso em atividades de pesquisa. Em particular, o projeto de pesquisa proposto demanda uma dedicação contínua e regular às várias atividades, o que justifica a adoção desse critério.

- Iniciativa: A resolução de problemas e dificuldades sem a interferência do orientador são exemplos de como a iniciativa pode se materializar durante o trabalho de iniciação científica.

- Habilidade técnica: O projeto proposto demanda entendimento profundo de algoritmos e de sua implementação, exigindo boa habilidade técnica dentro dessas áreas. A rapidez e correção na execução das tarefas planejadas são métricas a serem utilizadas para quantificar a habilidade técnica.

- Criatividade: O contato com o estado-da-arte na área do projeto e a busca de soluções inovadoras são ingredientes para o bolsista exercitar a sua criatividade, elemento fundamental para a formação de um pesquisador.

- Convivência: O trabalho em grupo e atividades de cooperação serão uma constante. A capacidade de interagir e construir resultados junto com outros colegas é um indício forte de sucesso nesse critério.

Capacidade crítica: O avanço da ciência se baseia pesadamente na avaliação por pares. Esse processo se inicia pela leitura e avaliação crítica de artigos técnicos, além da sua apresentação e discussão no âmbito do projeto. Em um segundo momento, será fomentada a participação e avaliação dos trabalhos dos pares, assim como a proposição de soluções. Finalmente, a elaboração de artigos para submissão demanda não só desenvolvimento da capacidade crítica como de síntese.

Expressão oral e escrita: A capacidade de escrita é fundamental para a publicação dos resultados de pesquisa e será exercitada no âmbito do projeto pela preparação de relatórios, realização de apresentações orais e elaboração de artigos para submissão.

Estes critérios serão avaliados durante todo o trabalho, e o resultado dessa avaliação será comunicado ao bolsista com o fim de auxiliá-lo a se aperfeiçoar.

## 5 Referencias

- [1] NEWMAN, M. E. Spread of epidemic disease on networks. *Physical review E*, APS, v. 66, n. 1, p. 016128, 2002.
- [2] TROTTIER, H.; PHILIPPE, P. Deterministic modeling of infectious diseases: theory and methods. *The Internet Journal of Infectious Diseases*, v. 1, n. 2, p. 3, 2001
- [3] KERMACK, W. O.; MCKENDRICK, A. G. A contribution to the mathematical theory of epidemics. In: THE ROYAL SOCIETY. *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*. [S.l.], 1927. v. 115, n. 772, p. 700–721.
- [4] BARABÁSI, A.-L. *Network science*. [S.l.]: Cambridge university press, 2016.
- [5] MITCHELL, M. *An introduction to genetic algorithms*. [S.l.]: MIT press, 1998
- [6] DARWIN, C.; WALLACE, A. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Zoological Journal of the Linnean Society*, Wiley Online Library, v. 3, n. 9, p. 45–62, 1858.