

Simulação da propagação da COVID-19 em municípios modelados por redes sociais a partir de registros telefônicos e dados censitários

Paulo Gabriel Nunes Cançado¹, Carolina Ribeiro Xavier¹,
Alexandre Gonçalves Evsukoff², Vinícius da Fonseca Vieira¹

¹Departamento de Ciência da Computação –
Universidade Federal de São João del-Rei (UFSJ)
São João del-Rei – MG – Brasil

²COPPE – Universidade Federal do Rio de Janeiro
Rio de Janeiro – RJ – Brasil

vinicius@ufsj.edu.br

Abstract. *This work presents a methodology for simulating the spread of COVID-19 in urban populations using a compartmental SIR model based on multilayer complex networks, combining Call Detail Records (CDR) and census data to build a network that captures various perspectives of social interaction, such as mobility, communication, family relationships, school/work ties, and religious activities. Through a Genetic Algorithm, it was possible to calibrate the model to real epidemiological bulletin data by defining the parameters that control the daily infection and recovery rates of individuals in four municipalities: Betim, Contagem, Juiz de Fora, and Uberlândia.*

Resumo. *Este trabalho apresenta uma metodologia para a simulação da propagação da COVID-19 em populações urbanas, utilizando um modelo compartimental SIR sobre redes complexas multicamadas, combinando dados de registros de chamadas telefônicas (Call Detail Records - CDR) e dados censitários para construir uma rede que captura diversas perspectivas de interação social, como mobilidade, comunicação, relações familiares, laços escolares/laborais e atividades religiosas. Através de um Algoritmo Genético foi possível ajustar o modelo a dados de boletins epidemiológicos reais, através da definição dos parâmetros controlam as taxas diárias de infecção e recuperação de indivíduos em quatro municípios: Betim, Contagem, Juiz de Fora e Uberlândia.*

1. Introdução

A rápida disseminação da COVID-19 [WHO 2020], doença infectocontagiosa responsável pela última pandemia na era contemporânea, deve-se, em grande parte, à facilidade de transmissão do vírus, que ocorre principalmente por gotículas de saliva ou secreções expelidas por pessoas infectadas. O contato com o vírus SARS-CoV-2 e a consequente contaminação foram especialmente críticos em populações mais vulneráveis dos pontos de vista econômico, demográfico ou social, que enfrentaram maiores dificuldades para adotar medidas preventivas eficazes [Chang et al. 2021]. A pandemia de COVID-19 foi oficialmente encerrada em 5 de maio de 2023, deixando impactos globais significativos. Foram registrados cerca de 700 milhões de casos e aproximadamente

7 milhões de mortes em todo o mundo [Worldometer 2025]. O Brasil foi o segundo país com maior número de óbitos registrados (708 mil), com cerca de 38 milhões de casos [Worldometer 2025], evidenciando o despreparo das instituições brasileiras diante da pandemia, dada a escassez de medidas eficazes para conter a propagação e evitar mortes.

Muitos modelos foram propostos para descrever a dinâmica de epidemias, deles ser classificados como modelos coletivos, associados a modelos homogêneos [Sarkar et al. 2020, Xavier et al. 2022], que partem da hipótese de que todos os indivíduos de uma população possuem as mesmas características e têm igual probabilidade de contato entre si. Entre eles, o modelo SIR (Suscetível-Infectado-Recuperado), proposto por [Kermack and McKendrick 1927], é um dos mais conhecidos e amplamente utilizados por sua simplicidade e capacidade de capturar os aspectos essenciais da disseminação de uma doença infecciosa. No contexto da COVID-19, modelos baseados no SIR foram empregados em diversos estudos [Costa et al. 2020, Xavier et al. 2022].

Ao assumir uma estrutura populacional homogênea, modelos coletivos podem não representar adequadamente a complexidade das interações sociais, os padrões de mobilidade e as desigualdades estruturais que influenciam diretamente a propagação do vírus. Isso motivou a extensão de modelos epidemiológicos clássicos para incorporar características topológicas intrínsecas às relações entre as pessoas, que definem redes sociais, como introduzido por [Pastor-Satorras et al. 2015] e utilizado em diversos outros estudos que utilizam modelos baseados em redes [Vespignani et al. 2015].

Neste trabalho, considera-se a simulação do espalhamento da COVID-19 através de um modelo compartimental SIR sobre uma rede social construída a partir de dados censitários e registros de ligações telefônicas (*Call Detail Records* - CDR), armazenados pelas companhias para possibilitar cobranças de maneira adequada entre diferentes operadoras. Dados CDR podem ser utilizados de maneira alternativa ao uso de questionários (*surveys*) combinados a novas tecnologias para armazenamento e processamento de grandes volumes de dados georreferenciados para o desenvolvimento de soluções para cidades baseadas em modelos computacionais centrados em dados, capazes de ajudar a caracterizar a organização geográfica e socioeconômica de uma população com grande resolução e aplicabilidade em contextos práticos [Lenormand and Ramasco 2016, Onnela et al. 2007]. Porém, registros telefônicos do tipo CDR podem introduzir vieses a partir da representatividade da população da cidade, dos estratos sociais, da região geográfica do método de extração e armazenamento dos dados, da duração da captura da base de dados, entre outros [Pestre et al. 2020]. Neste trabalho, o modelo de redes sociais incorpora aos dados CDR informações censitárias, a partir de dados do IBGE, para caracterizar a demografia da população e sua interação em diferentes perspectivas, utilizando abordagem proposta em um trabalho anterior ([Cançado 2025]). Através dos CDR, a comunicação entre indivíduos, registrada por dados digitais telefônicos, é utilizada como *proxy* de seus relacionamentos, e a posição em que as chamadas são realizadas é utilizada para inferir a mobilidade dos indivíduos em centros urbanos. Através de dados censitários, o modelo incorpora informações demográficas relacionadas à composição familiar, atividades escolares, de trabalho e religiosas. Essas perspectivas são representadas através de uma rede social projetada a partir de múltiplas camadas de interação [Battiston et al. 2014, Kivela et al. 2014], o que é bastante razoável para a modelagem do espalhamento de uma doença cujo contágio se dá através do contato em diferentes

dimensões. O modelo SIR baseado em redes multicamadas tem seus parâmetros ajustados por um Algoritmo Genético, que busca minimizar a diferença entre o resultado da simulação e a evolução do contágio observada em dados de boletins epidemiológicos reais, em quatro municípios do Estado de Minas Gerais, no Brasil: Betim, Contagem, Juiz de Fora e Uberlândia.

Neste contexto, este estudo tem como objetivo geral a definição de uma metodologia para a construção de modelos SIR (Suscetível-Infectado-Recuperado) para a simulação da propagação da COVID-19 em populações urbanas representadas por redes sociais multicamadas construídas sobre dados de registros telefônicos (CDR) e dados censitários. Para isso, são definidos os seguintes objetivos específicos:

- Construir uma rede multicamadas que represente aspectos sociais, de mobilidade e demográficos, combinando dados de CDR e dados censitários para capturar a movimentação de indivíduos, interações sociais e heterogeneidade estrutural;
- Simular a propagação da COVID-19 empregando o modelo epidemiológico Suscetível-Infectado-Recuperado (SIR) baseado em redes complexas e adaptado à topologia da rede multicamadas;
- Ajustar automaticamente os parâmetros β (taxa de infecção) e γ (taxa de recuperação) utilizando um Algoritmo Genético (AG), garantindo maior eficiência, reprodutibilidade e aplicabilidade do método em diferentes cenários e localidades.

2. Conceitos fundamentais

2.1. Modelo SIR (Suscetível-Infectado-Recuperado)

SIR (Suscetível-Infectado-Recuperado) é um modelo epidemiológico compartimental criado por Kermack e McKendrick [Kermack and McKendrick 1927] em 1927. Nesse modelo, a população é dividida em três compartimentos que representam os possíveis estágios em relação à doença. Matematicamente, o modelo pode ser descrito pelas Equações 1, 2 e 3.

$$\frac{\partial S}{\partial t} = -\beta S \frac{I}{N}, \quad (1)$$

A Equação 1 descreve a taxa de variação do número de indivíduos suscetíveis ao longo do tempo. Essa variação é negativa, pois os suscetíveis diminuem à medida que entram em contato com infectados. A fração $\frac{I}{N}$ representa a proporção de infectados na população total, e o parâmetro β quantifica a taxa de transmissão da doença, ou seja, a probabilidade de infecção por contato. Assim, o termo $\beta S \frac{I}{N}$ representa a taxa com que os suscetíveis se tornam infectados.

$$\frac{\partial I}{\partial t} = \beta S \frac{I}{N} - \gamma I \quad (2)$$

A Equação 2 representa a variação do número de infectados. O primeiro termo do lado direito, $\beta S \frac{I}{N}$, corresponde ao ingresso de novos infectados, como explicado anteriormente. Já o segundo termo, $-\gamma I$, descreve a recuperação dos indivíduos, sendo γ a taxa de recuperação, que determina a proporção de infectados que deixam esse compartimento por unidade de tempo.

$$\frac{\partial R}{\partial t} = \gamma I \quad (3)$$

A Equação 3 expressa o crescimento do número de indivíduos recuperados. Ela é proporcional à quantidade de infectados no tempo e regida pela mesma taxa γ , assumindo

que os recuperados adquirem imunidade permanente ou são removidos da população. As variáveis S , I e R representam, respectivamente, a quantidade de indivíduos suscetíveis, infectados e recuperados. Além disso, considera-se que a população total permanece constante, sendo representada pela Equação 4:

$$N = S + I + R \quad (4)$$

Os parâmetros β e γ , introduzidos nas equações diferenciais do SIR, descrevem a dinâmica de propagação da doença no modelo SIR. O parâmetro β representa a taxa de infecção, isto é, a probabilidade de transmissão em um contato entre um indivíduo suscetível e um infectado. Já γ indica a taxa de recuperação, correspondendo à fração de infectados que se recuperam por unidade de tempo. Ambos os parâmetros variam conforme a doença e devem ser estimados a partir de dados observacionais. Os parâmetros β e γ possibilitam calcular o número de reprodução básica (R_0) e pode ser calculado como:

$$R_0 = \frac{\beta}{\gamma}. \quad (5)$$

2.1.1. Modelo SIR em Redes Complexas

Em modelos compartimentais clássicos, como o SIR, na forma como foi proposto em [Kermack and McKendrick 1927], assume-se que todos os indivíduos têm a mesma probabilidade de serem infectados, de se recuperar ou de propagar a doença, sem considerar diferenças nas interações sociais, comportamentais ou nas características individuais [Keeling and Rohani 2008]. Assim, desconsidera-se a diversidade na forma como os indivíduos interagem entre si, o que pode ser fundamental para a compreensão da propagação de uma doença contagiosa, como é o caso da COVID-19. Modelos baseados em redes buscam representar a complexidade das interações, utilizando informações mais detalhadas sobre topologia do contato entre as pessoas, a intensidade das interações, a distribuição geográfica e a estrutura social, permitindo uma representação mais precisa da propagação da doença [Keeling and Rohani 2008]. Pode-se, então, utilizar o arcabouço metodológico do estudo de ciência de redes e teoria dos grafos para modelar essas interações, proporcionando uma simulação mais realista dos padrões de transmissão, que varia conforme o comportamento e a posição dos indivíduos na rede.

Neste trabalho, considera-se o modelo SIR baseado em Redes Complexas, como proposto por Pastor-Satorras e Vespignani [Pastor-Satorras et al. 2015], que busca representar a natureza probabilística da propagação da doença na população de forma heterogênea. Assim como no modelo SIR clássico, os dois parâmetros probabilísticos, β e γ , representam, respectivamente, a probabilidade de um nó infectado contaminar um nó suscetível na rede, desde que exista uma aresta entre eles e a probabilidade de um nó infectado se recuperar, passando para o estado de recuperado. Como a menor porção de tempo simulada neste trabalho é de um dia, os parâmetros encontrados referem-se às probabilidades diárias de infecção e recuperação. Uma explicação mais detalhada sobre o assunto pode ser encontrada em [Pastor-Satorras et al. 2015].

2.2. Algoritmos Genéticos

Algoritmos Genéticos (AGs) são heurísticas de busca e otimização [Mitchell 1998] inspiradas na teoria da evolução de Charles Darwin que funcionam a partir de uma “população” de soluções potenciais para um problema, representadas como cromossomos. Através de um processo iterativo, as soluções são avaliadas por uma função de aptidão (*fitness*) e as melhores são selecionadas para “reproduzir” para uma nova iteração, combinando suas características através de *crossover* (recombinação) e sofrendo mutações ocasionais. Esse ciclo permite que o algoritmo explore o espaço de busca e evolua em direção a soluções cada vez mais adequadas ao longo das gerações.

Nos Algoritmos Genéticos os indivíduos (soluções potenciais) mais adaptados têm mais chance de sobreviver e uma nova população é formada com base nos indivíduos potencialmente mais aptos da população anterior. Consequentemente, as soluções encontradas pelo algoritmo tendem a melhorar a cada nova geração de indivíduos.

3. Materiais e métodos

3.1. Modelagem da rede social

Uma das maiores preocupações em aplicações que utilizam modelos baseados em redes sociais [Onnela et al. 2007, Ribeiro et al. 2021] construídos a partir de dados CDR é evitar que a rede social introduza vieses que façam com que, de fato, não reflitam interações entre pessoas [Vieira et al. 2023, Scabini et al. 2021]. Essa questão é tratada aqui com a integração de dados censitários para a definição de diferentes perspectivas de interação entre indivíduos, representadas por diferentes camadas na rede construídas com base na distribuição demográfica da população em diferentes dimensões através de uma metodologia proposta em [Cançado 2025].

Para isso, tomamos como base o formalismo introduzido por Battiston *et al.* [Battiston et al. 2014] e definimos, para cada cidade específica, uma rede multicamadas, que pode ser representada por um grafo $G^{[\alpha]}$, composto de um conjunto de vértices $v_i \in V$, tal que $|V| = n$, que representam os indivíduos de uma determinada cidade, e l camadas. Cada camada α ($\alpha = [1, \dots, l]$) representa a relação entre os indivíduos v_i sob uma perspectiva α (ex.: comunicação, mobilidade, família) e pode, então, ser representada por $G^{[\alpha]} = (V, E^{[\alpha]})$. $E^{[\alpha]}$ é o conjunto de arestas (v_i, v_j) que representa as relações entre os indivíduos v_i e v_j sob uma perspectiva α .

A metodologia proposta é aplicada a quatro municípios, localizados no estado de Minas Gerais, no Brasil: Betim, Contagem, Juiz de Fora e Uberlândia. Define-se, então, diferentes camadas, que representam as relações dos indivíduos: baseadas em comunicação ($\alpha = 1$), mobilidade ($\alpha = 2$), familiares ($\alpha = 3$), escolares/laborais ($\alpha = 4$), religiosas ($\alpha = 5$) e aleatórias ($\alpha = 6$).

Para a construção das camadas baseadas em CDR ($\alpha = 1$ e $\alpha = 2$), são utilizados dados de registros telefônicos realizadas nos municípios de interesse em um período de 30 dias entre 15/03/2013 e 14/04/2013¹. Dados CDR geralmente armazenam informação de um ID dos usuários de uma chamada, assim como sua duração, assim como o ID

¹Os dados CDR foram obtidos através de um convênio entre o grupo de pesquisa dos autores do trabalho e uma grande operadora de telefonia celular do Brasil. Todos os dados foram anonimizados para a condução do trabalho.

da antena utilizada pelo dispositivo móvel para realizar a chamada o que, juntamente com sua localização, fornece uma estimativa a localização do usuário. Na camada de interações sociais baseadas na comunicação ($G^{[1]} = (V, E^{[1]})$), as relações entre os indivíduos i e j são representadas como uma aresta $(v_i, v_j) \in E^{[1]}$ quando são observadas chamadas telefônicas entre ambos, obedecendo alguns critérios, como adotado em outros trabalhos utilizados como referência [Lenormand and Ramasco 2016]. A mobilidade dos indivíduos é incorporada ao modelo de rede proposto ($G^{[2]} = (V, E^{[2]})$), com o objetivo de representar encontros que podem ocorrer, de maneira deliberada ou não, entre pessoas que se deslocam na cidade. Isso é feito a partir da construção de uma matriz Origem-Destino, seguindo a metodologia proposta por [Chaves et al. 2023], que utiliza as posições das antenas das chamadas telefônicas para inferir seu deslocamento no espaço urbano.

O modelo multicamadas proposto considera também dados censitários, com a finalidade de reduzir vieses na distribuição demográfica que pode ocorrer na construção de modelos baseados puramente em dados CDR. Para isso, é utilizada uma estratégia introduzida pelo trabalho de Scabini *et al.* [Scabini et al. 2021] e os dados censitários são utilizados para determinar parâmetros $\mu^{[\alpha]}$, que controlam as probabilidades de existência de arestas em cada uma das camadas α , que estimam as relações nas perspectivas familiares ($G^{[3]} = (V, E^{[3]})$), escolares/laborais ($G^{[4]} = (V, E^{[4]})$), religiosas ($G^{[5]} = (V, E^{[5]})$) e aleatórias ($G^{[6]} = (V, E^{[6]})$). As camadas baseadas em dados censitários no modelo proposto neste trabalho são construídas a partir de dados públicos disponibilizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE)². Especificamente, foram considerados os dados do Censo 2010, para preservar a coerência com os dados CDR disponíveis para o trabalho.

Após a construção das multicamadas, o modelo é agregado em uma rede resultante $G = (V, E)$ que tem, para cada aresta $(v_i, v_j) \in E$, uma ponderação $w(i, j) = \sum_{\alpha=1}^l w^{[\alpha]}$, $\alpha = [1, \dots, l]$. Dessa forma, indivíduos que compartilham múltiplos contextos sociais têm suas interações representadas por conexões mais intensas do que aqueles que interagem apenas em um único contexto.

3.2. Dados epidemiológicos

O ajuste de parâmetros do modelo SIR, através de Algoritmo Genético, é realizado sobre dados de boletins epidemiológicos disponibilizados pelo Ministério da Saúde³. Nesses boletins, são divulgados os números diários de novas infecções, mortes e a quantidade atual de indivíduos recuperados. Nessa etapa, é necessário ajustar os dados de acordo com o número de nós da rede. Para isso, determina-se a fração entre a quantidade de nós e a população total da cidade, aplicando em cada valor diário.

3.3. Algoritmo Genético

O Algoritmo Genético desenvolvido visa encontrar os valores dos parâmetros β (taxa de infecção) e γ (taxa de recuperação) que resultem em simulações da propagação do vírus na rede capazes de reproduzir, com a maior fidelidade possível, o comportamento observado nos dados reais. A Função Objetivo (FO) é representada por:

$$\min f(s) = \theta, 7 \sum_{i=1}^D (IR_i - IS_i)^2 + \theta, 3 \sum_{i=1}^D (RR_i - RS_i)^2 \quad (6)$$

²<https://www.ibge.gov.br/>, acessado em 15 de maio de 2026.

³<https://covid.saude.gov.br/>, acessado em 15 de maio de 2026.

onde IR e RR correspondem aos números observados de infectados e recuperados, enquanto IS e RS são os valores simulados. D indica o número de dias simulados. Os pesos 0,7 e 0,3 na Equação 6 atribuem maior importância à redução do erro associado aos infectados, considerados mais relevantes para a análise.

Todas as simulações partem do mesmo conjunto inicial de nós infectados e recuperados, garantindo consistência nos experimentos.

A cada geração, os indivíduos de uma população são selecionados em pares para gerar a próxima população, composta de 10 indivíduos por meio da técnica de roleta viciada, onde a probabilidade de escolha de um indivíduo é inversamente proporcional à sua nota. A recombinação dos pais ocorre por meio do *crossover* de um único ponto. Os indivíduos gerados passam por um processo de mutação, que permite a exploração de novas soluções e evita a convergência prematura para ótimos locais. Além disso, é aplicado o mecanismo de elitismo, no qual uma porcentagem fixa dos melhores indivíduos é copiada diretamente de uma geração para a próxima, garantindo que as soluções mais aptas não sejam perdidas durante o processo evolutivo, que dura até 20 gerações. Testes empíricos indicaram que uma taxa de mutação de 15% e a preservação de 10% a 40% da elite da população são adequadas para este problema.

4. Experimentos e discussão

Para cada uma das quatro cidades estudadas, foi selecionado um intervalo em que a curva de infectados apresenta um comportamento razoavelmente estável, sem inversões bruscas de tendência, em um intervalo que varia de 21 a 34 dias. Dessa maneira, o modelo assume valores constantes para os parâmetros β e γ ao longo de toda a simulação.

Os resultados obtidos para a melhor simulação em cada uma das cidades estudadas, considerando os parâmetros descritos na Seção 3, são apresentados na Tabela 1.

Tabela 1. Resultados dos ajustes de parâmetros para cada cidade.

Cidade	Beta	Gamma	MSE Beta	MSE Gamma	R2 Inf. (%)	R2 Rec. (%)
Betim	0.048	0.081	75	202	0.90	0.97
Contagem	0.016	0.006	1668	1801	0.85	0
Juiz de Fora	0.204	0.395	421	11662	0.44	0.91
Uberlândia	0.056	0.047	8326	7820	0.94	0.99

Os resultados apresentados na Tabela 1 mostram, de modo geral, um bom ajuste aos dados observados, o que pode ser corroborado com a observação das Figuras 1 e 2, que ilustram, respectivamente, as curvas de infectados e recuperados, para os dados reais (linha sólida azul) e os dados simulados (linha tracejada laranja).

Mesmo percebendo que as curvas simuladas de Infectados e Recuperados são capazes de capturar, em grande parte do intervalo, as tendências das curvas dos dados originais, o presente trabalho define um arcabouço metodológico que pode ser ajustado para diferentes cenários de modelagem de doenças contagiosas, mas que precisa ter suas limitações consideradas e adaptadas a situações específicas. Primeiramente, é impossível desconsiderar que a construção das redes insere um importante viés que é propagado no processo de ajuste. Por exemplo, na cidade de Juiz de Fora, cuja rede apresentou um nível de conectividade significativamente superior ao das demais, característica que pode

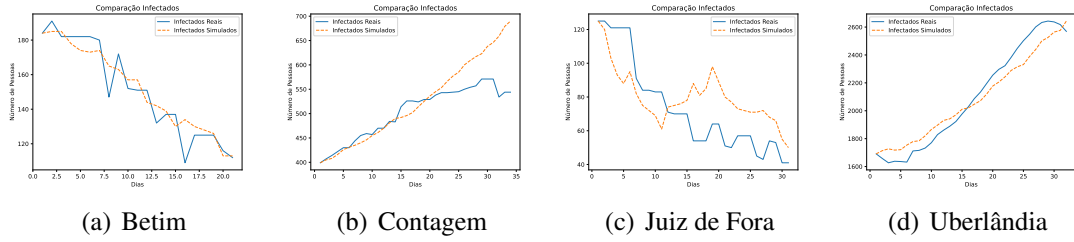


Figura 1. Indivíduos infectados (reais e simulados) para as cidades analisadas.

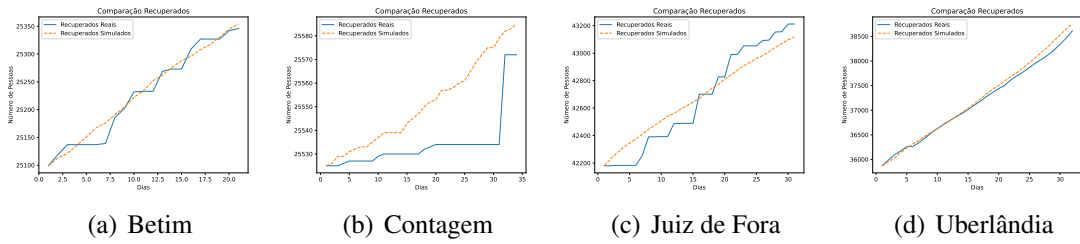


Figura 2. Indivíduos recuperados (reais e simulados) para as cidades analisadas.

estar associada aos dados de CDRs, mais abundantes naquela região. À medida em que a rede gerada se distancia da representatividade do contexto social observado em um determinado ambiente urbano, mais dificuldade o modelo enfrenta para ajustar os parâmetros β e γ de forma eficaz. Essa desconexão entre a topologia da rede e o comportamento real dos contatos pode explicar, ao menos em parte, a piora de desempenho das simulações em alguns casos, como o de Juiz de Fora.

Vale destacar também que os próprios dados epidemiológicos utilizados para calibrar o modelo constituem uma fonte relevante de limitação e não é incomum que apresentem inconsistências que dificultam o ajuste. Em Juiz de Fora, por exemplo, há uma distribuição irregular dos novos casos, com registros concentrados em dias específicos da semana, o que resulta em curvas de infectados reais com variações acentuadas. De modo semelhante, na cidade de Contagem, verificou-se que os recuperados foram contabilizados de forma acumulada em determinados dias. Esse padrão gerou saltos abruptos na curva de recuperados reais. Esses ruídos tornam as curvas irregulares, dificultando a identificação de tendências claras e, consequentemente, a obtenção de simulações que se ajustem bem aos dados observados. Esses fatores, quando somados, revelam os desafios enfrentados pelo modelo diante de dados e estruturas de rede que não representam fielmente a realidade da dinâmica epidêmica local.

5. Conclusões e trabalhos futuros

Este trabalho apresentou e avaliou uma metodologia para a simulação da propagação da COVID-19 em populações urbanas, utilizando um modelo compartimental SIR fundamentado em redes complexas multicamadas. A abordagem proposta combina dados de registros de chamadas telefônicas (CDR) e dados censitários do IBGE para construir uma rede que captura diversas perspectivas de interação social, como mobilidade, comunicação, relações familiares, laços escolares/laborais e atividades religiosas.

A pesquisa evidenciou desafios importantes na modelagem computacional de epi-

demias baseada em dados reais. Verificou-se que a construção da rede pode introduzir vieses significativos e, além disso, inconsistências nos boletins epidemiológicos reais constituíram um forte obstáculo. Por fim, o modelo assume, em sua configuração atual, que os parâmetros β e γ permanecem constantes ao longo de todo o período simulado, limitando sua capacidade de refletir mudanças na dinâmica de contágio.

A utilização de um Algoritmo Genético demonstrou ser uma estratégia eficaz para o ajuste dos parâmetros do modelo. Através da minimização do erro em relação aos dados reais disponibilizados pelo Ministério da Saúde, o algoritmo foi capaz de encontrar taxas diárias de infecção (β) e recuperação (γ) que geraram simulações consistentes. De maneira geral, as curvas simuladas conseguiram capturar as tendências de comportamento das curvas de indivíduos infectados e recuperados em quatro municípios analisados (Betim, Contagem, Juiz de Fora e Uberlândia).

Considerando o arcabouço metodológico estabelecido, propõe-se, para trabalhos futuros, adaptar o modelo SIR para permitir que os parâmetros β e γ variem ao longo do tempo, superando a limitação atual de valores constantes e permitindo refletir eventos externos, como a implementação de medidas de distanciamento social, surgimento de novas variantes ou campanhas de vacinação. Além disso, pretende-se aprimorar as técnicas de construção das multicamadas baseadas em CDR, desenvolvendo métodos para equilibrar a conectividade em regiões com volume desproporcional de dados telefônicos.

Agradecimentos

Este trabalho contou com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Referências

- Battiston, F., Nicosia, V., and Latora, V. (2014). Structural measures for multiplex networks. *Phys. Rev. E*, 89:032804.
- Cançado, P. G. N. (2025). Ajuste de um modelo epidemiológico em redes sociais definidas por dados de múltiplas fontes. Dissertação (mestrado em ciência da computação), Universidade Federal de São João del-Rei (UFRJ), São João del-Rei, MG, Brasil.
- Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., and Leskovec, J. (2021). Mobility network models of covid-19 explain inequities and inform reopening. *Nature*, 589(7840):82–87.
- Chaves, J. C., da Silva, M. A., de Souza Alencar, R., Evsukoff, A. G., and da Fonseca Vieira, V. (2023). Human mobility and socioeconomic datasets of the rio de janeiro metropolitan area. *Data in Brief*, 51:109695.
- Costa, G. S., Cota, W., and Ferreira, S. C. (2020). Outbreak diversity in epidemic waves propagating through distinct geographical scales. *Physical Review Research*, 2(4):043306.
- Keeling, M. J. and Rohani, P. (2008). *Modeling infectious diseases in humans and animals*. Princeton university press.

- Kermack, W. O. and McKendrick, A. G. (1927). 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*, 115(772):700–721.
- Kivela, M., Arenas, A., Barthélemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, 2(3):203–271.
- Lenormand, M. and Ramasco, J. J. (2016). Towards a better understanding of cities using mobility data. *Built Environment*, 42:356–364(9).
- Mitchell, M. (1998). *An introduction to genetic algorithms*. MIT press.
- Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., and Barabási, A.-L. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the national academy of sciences*, 104(18):7332–7336.
- Pastor-Satorras, R., Castellano, C., Van Mieghem, P., and Vespignani, A. (2015). Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925.
- Pestre, G., Letouzé, E., and Zagheni, E. (2020). The ABCDE of Big Data: Assessing Biases in Call-Detail Records for Development Estimates. *The World Bank Economic Review*, 34(1):89–97.
- Ribeiro, J. M., Alencar, R., Martins, G., Xavier, C., Evsukoff, A., and Vieira, V. (2021). Caracterização da relação entre redes sociais e mobilidade de indivíduos em contextos urbanos. In *Anais do X Brazilian Workshop on Social Network Analysis and Mining*, pages 32–43. SBC.
- Sarkar, K., Khajanchi, S., and Nieto, J. J. (2020). Modeling and forecasting the covid-19 pandemic in india. *Chaos, Solitons & Fractals*, 139:110049.
- Scabini, L. F., Ribas, L. C., Neiva, M. B., Junior, A. G., Farfán, A. J., and Bruno, O. M. (2021). Social interaction layers in complex networks for the dynamical epidemic modeling of covid-19 in brazil. *Phys. A: Stat. Mech. and its App.*, 564:125498.
- Vespignani, A., Pastor-Satorras, R., Van Mieghem, M., and Castellano, C. (2015). Epidemic processes in complex networks. *Rev. Mod. Phys.*, 87(925).
- Vieira, V. d. F., Alencar, R., Evsukoff, A. G., and Samaniego, H. (2023). *Proceedings XXIV GEOINFO*, page 220–231.
- WHO (2020). Coronavirus disease 2019 (covid-19): Situation report. Situation report 51, World Health Organization. Acesso em: 08 jan. 2024.
- Worldometer (2025). Coronavirus cases. <https://www.worldometers.info/coronavirus/>. Acesso em: 08 jan. 2024.
- Xavier, C. R., Oliveira, R. S., Vieira, V. F., Lobosco, M., and dos Santos, R. W. (2022). Characterisation of omicron variant during covid-19 pandemic and the impact of vaccination, transmission rate, mortality, and reinfection in south africa, germany, and brazil. *BioTech*, 11(2).