

# Rastreamento de Contatos em Modelo de Espalhamento de Doenças Infecciosas em Redes Complexas em Tempo Discreto

J. H. N. Batista

Setembro 2022

## Resumo:

A transmissão acelerada de doenças contagiosas são umas das principais ameaças a saúde global. Visando contribuir com a atenuação do espalhamento desses males, o rastreamento de contato e a identificação de possíveis infectados é crucial. Nesse sentido, modelamos o espalhamento de uma doença infecciosa em um grafo estático (onde os vértices são os membros da população) em tempo discreto usando um modelo SIR simplificado. Esse trabalho propõe um algoritmo de Monte Carlo que estima a probabilidade de um vértice ter sido infectado, condicionado a informação sobre a infecção de outros vértices.

## 1 Introdução

As doenças infecciosas são um grande risco para a saúde global, como mostrou a recente pandemia de COVID-19. Além disso, quando a OMS compilou uma lista das 10 principais ameaças à saúde global para se concentrar em 2019, os desafios da lista - que continham uma coleção de questões sérias, desde mudanças climáticas até instalações de saúde inadequadas - eram, mais da metade, constituídos por desafios ligados a doenças infecciosas, tanto emergentes quanto históricas.

Em 1927, Kermack e McKendrick [1] deram uma contribuição seminal para a compreensão da propagação de epidemias, introduzindo um modelo matemático agora conhecido como modelo de compartimentos. Desde então, modelos matemáticos, vêm sendo uma ferramenta útil para revelar a dinâmica de transmissão de doenças infecciosas. Nos últimos 3 anos esses modelos foram amplamente usados no estudo da COVID-19 [2, 3, 4].

Os modelos chamados de *mean-field* (campo médio), normalmente baseados em equações diferenciais, supõem que a população seja homogênea e são ótimas ferramentas para o estudo de fenômenos *macroscópicos*, como predição do número de infectados e mortos [3, 4] e formulação de estratégias de vacinação [5], testagem [2] e distanciamento social [6].

Contudo, tais modelos não são capazes de prevê fenômenos em menor escala ou considerar a heterogeneidade da população. Nesse contexto surgem os modelos de espalhamento baseados em grafos, que são capazes de estudar tais fenômenos.

Sabe-se que muitas doenças infecciosas, como é o caso da COVID-19, apresentam altas taxas de pacientes assintomáticos, o que favorece o espalhamento da doença, pois tais pacientes não se isolam por acharem que não estão contaminando a população [7]. Desse modo, o rastreamento de contato é ferramenta fundamental no combate a uma pandemia, é preciso identificar possíveis pacientes assintomáticos, realizar testagem e sugerir o isolamento social. Dada a importância desse ato, muitos trabalhos surgiram tentando mapear os contatos entre os indivíduos usando dados de geolocalização, wifi e bluetooth [8, 9, 10].

Nesse trabalho, foi utilizado um modelo SIR teórico em tempo discreto, de uma doença que se espalha em um grafo (estático). Iremos supor que, para alguns nós da rede que foram infectados, conhecemos se eles foram infectados, mas para outros nós infectados, nós não conhecemos se eles foram ou não infectados. Assumimos também que para os nós que sabemos se foram infectados, sabemos também o dia da infecção. O objetivo é elaborar um método estatístico, baseado na técnica de Monte Carlo (MC) e *Importance Sampling* [11], para estimar a probabilidade de um dado vértice  $v$  ter sido infectado, condicionado aos dados observados.

## 2 Revisão Bibliográfica

Em [12] o autor faz um modelo de espalhamento de doenças do tipo SIS em que cada nó do grafo é uma região (cidade, bairro) que possui uma proporção  $x_i(t)$  de infectados, onde  $i$  é um vértice do grafo e  $t$  o instante de

tempo (aqui admite-se tempo discreto), tais proporções interagem entre si por meio de uma matriz de adjacência contendo o peso da interação entre cada vértice; o artigo visa resolver um problema de controle ótimo da melhor estratégia para reduzir o efeito da doença, a partir de medidas de isolamentos sociais intra e inter as regiões. Já em [13] é estudado um modelo, em tempo contínuo, de transmissão de doença do tipo SEIR, onde para cada nó  $v$  do grafo (que também representa um região) está associado as quantidades  $s_v(t)$ ,  $e_v(t)$ ,  $i_v(t)$  e  $r_v(t)$ , respectivamente, proporção de suscetíveis, exposto, infectados e recuperados, tais quantidade interagem entre si por meio de uma matriz de adjacências.

Os modelos a cima não consideram o contato *face-a-face* dos indivíduos da população, i.e., os vértices do grafo não modelam diretamente as pessoas. Mas na literatura existem muitos trabalhos nesta direção. Tais modelos podem ser estudados considerando um modelo markoviano em tempo contínuo [14] ou em tempo discreto [15, 16, 17, 18, 19]. Esse trabalho terá foco no estudo em tempo discreto.

Em [15] o autor faz um estudo considerando o grafo da conexão entre as pessoas como sendo um Lattice  $D$ -dimensional, i.e., um grafo  $G = (V, E)$ , onde

$$V = (\mathbb{Z} \cap [-n, n])^D, E = \{(v, u) : \|u - v\| = 1, u, v \in V\}.$$

Nesse modelo são considerados alguns vértices “desocupados”, a alocação dos vértices ocupados (que muda a cada instante de tempo) é dada aleatoriamente de forma semelhante ao modelo de Ising (amplamente estudado em mecânica estatística) levando em consideração a distribuição das idades, enquanto que a transmissão da doença se assemelha a um modelo SIR; como o grafo das conexões se altera, este é um exemplo de um modelo com grafo dinâmico. Em [16] o autor faz um híbrido entre o modelo anterior e um modelo de grafo em cidades, nesse artigo, a dinâmica em cada cidade se dá por um modelo de Lattice, mas além disso, pode haver contágio entre cidades vizinhas. Em ambos os trabalhos, as simulações são feitas usando o método de Monte Carlo.

Na literatura, técnicas de espalhamentos em grafos vão muito além da epidemiologia, sendo comumente usada no estudo de espalhamento de rumores, informação e fakenews [20, 21].

Nesse trabalho, consideraremos um grafo estático, tal como em [19], i.e., o grafo é o mesmo em todo o tempo. Em [19] é construído um grafo onde os vértices representam pessoas de uma fração da população da Arábia Saudita, usando dados reais. Nesse trabalho as arestas que ligam os vértices são de 3 tipos: familiar (ocorrem quando os indivíduos são da mesma família), social (ocorrem quando indivíduos tem propriedades parecidas - moram perto, mesma idade, mesmo sexo etc) e aleatórias (simulam o contato com pessoas quaisquer, taxista, vendedores etc) - de modo que esse grafo possui a propriedade de *small world* - e simulam um modelo em tempo discreto da disseminação da COVID-19 (semelhante a um modelo SIRS) sendo cada instante de tempo equivalente a uma dia, via o método MC. Os resultados são muito próximo a realidade nos dois primeiros meses da pandemia. Essa simulação é usada para estimar o efeito de medidas de contenção específicas, como fechamento de escolas.

Em [18] o autor considera um grafo gerado aleatoriamente com a propriedade de ser *livre de escala*, e cuja a distribuição dos graus segue uma lei de potência,  $P(k) \sim k^{-\gamma}$ . O modelo de espalhamento da doença é um modelo SIS. É considerado um grafo  $G = (V, E)$  - com  $|V| = N$  -, associado a uma matriz de pesos  $R = [r_{ij}]_{ij}$ , onde  $r_{ij} > 0$  somente quando os vértices  $i$  e  $j$  estão conectados.  $G$  representa a ligação face-a-face dos indivíduos, enquanto  $r_{ij}$  é igual a probabilidade do nó  $i$  infectar o nó  $j$  no próximo instante de tempo, dado que  $i$  está infectado no tempo atual. Um nó infectado retorna a ser suscetível com probabilidade  $\mu$  para cada unidade de tempo. A dinâmica segue da seguinte forma

- No tempo inicial  $t = 0$  uma fração da população começa infectada;
- Em um tempo  $t$ , cada nó  $j$  que é vizinho de um nó  $i$  infectado possui probabilidade  $r_{ij}$  de ser infectado no tempo  $t + 1$ ;
- Após isso, cada nó infectado no tempo  $t$  retorna a ser suscetível no tempo  $t + 1$  com probabilidade  $\mu$ .

Nesse artigo é estudado a influência dos parâmetros na *densidade esperada de infecção*, que pode tanto ser calculada tanto analiticamente quanto usando MC, o artigo compara os dois métodos e conclui que o método de MC fornece uma boa estimativa.

No que diz respeito a inferência de parâmetros em modelos baseados em grafos, em [17] o autor encontra o estimador de máxima verossimilhança do vértice inicial (paciente 0) de um modelo de espalhamento de doença em grafos, em um modelo em que alguns dos vértices infectados são conhecidos. Muitos trabalhos na literatura utilizam o método de MC para estimar propriedades de grafos como em [22], mesmo para trabalhos não relacionados a contágio.

### 3 Metodologia e Modelo

O modelo proposto é principalmente baseado em [18, 19]. Consideraremos um grafo  $G = (V, E)$ , com  $|V| = N$  a qual está associado uma matriz e uma matriz  $R = [r_{ij}]_{ij}$  onde  $r_{ij} > 0$  somente quando  $i$  e  $j$  estão conectados. O modelo será do tipo *SIR*, onde vamos assumir que, uma vez que uma pessoa infectada se recupere, ela não se infectará novamente (o que é uma boa aproximação para curtos períodos de tempo). Aqui, vamos fixa que a taxa de recuperação definida em [18] é dada por  $\mu = 1$ , fazemos isso para simplificar a estrutura da árvore de infecção da dinâmica, tal pressuposto significa que consideramos que o tempo em que cada pessoa permanece infectada é constante. Essa modelo, pode também ser usado para propagação de rumor. Vamos supor conhecido o paciente 0. A dinâmica do modelo seguirá da seguinte forma,

- Inicialmente, no tempo  $t = 0$ , um pessoa está infectada.
- No tempo  $t$ , cada vértice infectado “tentará” infectar os vértices vizinhos que são suscetíveis com probabilidade de sucesso igual a  $r_{ij}$ . No tempo  $t + 1$  todos os vértice infectados no tempo  $t$  tornam-se recuperados e os novos vértices infectados continuam a dinâmica.

Considere que avaliamos o processo até um tempo  $T$ . Assuma que para um conjunto  $\{v_1, v_2, \dots, v_m\} = V' \subset V$  de nós que foram infectados, é conhecido o instante da infecção  $t_1, t_2, \dots, t_m$ . Chamamos de  $C$  o evento,

$$C = \{v_i \text{ foi infectado no tempo } t_i, i = 1, 2, \dots, m\}.$$

Dado  $v \in V$ , será apresentado um algoritmo que estima

$$\mathbb{P}(v \text{ ter sido infectado} | C).$$

Para testar o algoritmo proposto, foram geradas redes aleatórias de 4 tipos diferentes:

- Grafo de Erdős-Rényi [23, 24];
- Grafo de Barabási-Albert (Livre de Escalas) [25, 26];
- Grafo de Watts-Strogatz (Small World) [27, 28];
- Lattice  $D$ -dimensional (com fronteira periódica).

Os três primeiros desses modelos de geração de grafos são úteis por possuírem características comumente presente em redes do mundo real. O último não é tão fidedigno, mas é útil para entender a limitação do algoritmo, pois é um grafo pouco conexo o que possivelmente torna-se um desafio.

Após gerar a estrutura do grafo, adicionaremos pesos aleatoriamente a cada uma das arestas.

Para cada um dos 4 grafos, geraremos 3 diferentes realizações do processo de infecção até um tempo  $T$  pré-fixado. Em seguida, mediremos a eficiência do algoritmo proposta em cada um dos 12 grafos, cada um deles com a mesma quantidade de vértices  $N$ .

Para gerar os grafos tomou-se  $N = ?$ ,  $T = ?$  e

- Probabilidade  $p = 0.036$  de inserção de aresta, no grafo de Erdős-Rényi;
- Número de arestas  $m = 0.018N(N - 1)$  para o modelo de Barabási-Albert;
- Com  $k = 0.072(N - 1)$  arestas ligadas em cada nó da rede para o modelo de Watts-Strogatz;
- $D = 4$  no modelo de Lattice.

A justificativa dessa escolha é manter a densidade de aresta,  $\frac{|E|}{\binom{N}{2}}$ , próximo de 0.036 que é o observado em [19], a menos para o modelo do Lattice, que foi escolhido  $D$  para replicar [15].

### 4 Observação:

(Não estabeleci nessa seção o horizonte  $T$  da pandemia, nem o número de vértices  $N$  e nem os parâmetros para gerar os pesos. Completarei essa parte da metodologia somente na A2, pois preciso fazer teste para definir isso)

## References

- [1] William Ogilvy Kermack and Anderson G McKendrick. 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, 1927.
- [2] Marcos Amaku, Dimas Tadeu Covas, Francisco Antonio Bezerra Coutinho, Raymundo Soares Azevedo Neto, Claudio Struchiner, Annelies Wilder-Smith, and Eduardo Massad. Modelling the test, trace and quarantine strategy to control the covid-19 epidemic in the state of são paulo, brazil. *Infectious Disease Modelling*, 6:46–55, 2021.
- [3] Punam Bedi, Shivani, Pushkar Gole, Neha Gupta, and Vinita Jindal. Projections for covid-19 spread in india and its worst affected five states using the modified seird and lstm models, 2020.
- [4] Seema Patrikar, Deepti Poojary, DR Basannar, DS Faujdar, and Renuka Kunte. Projections for novel coronavirus (covid-19) and evaluation of epidemic response strategies for india. *Medical Journal Armed Forces India*, 76(3):268–275, 2020.
- [5] Boris Shulgin, Lewi Stone, and Zvia Agur. Pulse vaccination strategy in the sir epidemic model. *Bulletin of mathematical biology*, 60(6):1123–1148, 1998.
- [6] M. A. Ridenti, L. K. Teles, A. Maranhão, and V. K. Teles. Mathematical modeling and investigation on the role of demography and contact patterns in social distancing measures effectiveness in covid-19 dissemination, 2022.
- [7] Zhiru Gao, Yinghui Xu, Chao Sun, Xu Wang, Ye Guo, Shi Qiu, and Kewei Ma. A systematic review of asymptomatic infections with covid-19. *Journal of Microbiology, Immunology and Infection*, 54(1):12–16, 2021.
- [8] Higor S Monteiro, Shaojun Luo, Saulo DS Reis, Carles Igual, Antonio S Lima Neto, Matias Travizan, Jose Soares De Andrade Jr, Hernan Makse, et al. Superspreading k-cores at the center of covid-19 pandemic persistence. *Bulletin of the American Physical Society*, 66, 2021.
- [9] Sadamori Kojaku, Laurent Hébert-Dufresne, Enys Mones, Sune Lehmann, and Yong-Yeol Ahn. The effectiveness of backward contact tracing in networks. *Nature physics*, 17(5):652–658, 2021.
- [10] Marco Mancastroppa, Andrea Guizzo, Claudio Castellano, Alessandro Vezzani, and Raffaella Burioni. Side-ward contact tracing and the control of epidemics in large gatherings. *Journal of the Royal Society Interface*, 19(190):20220048, 2022.
- [11] Gino Biondini. An introduction to rare event simulation and importance sampling. In *Handbook of Statistics*, volume 33, pages 29–68. Elsevier, 2015.
- [12] Carlo Cenedese, Lorenzo Zino, Michele Cucuzzella, and Ming Cao. Optimal policy design to mitigate epidemics on networks using an sis model, 2021.
- [13] Giovanni Naldi and Giuseppe Patane’. A graph-based modelling of epidemics]A Graph-Based Modelling of Epidemics: Properties, Simulation, and Continuum Limit, publisher = arXiv, year = 2022, copyright = Creative Commons Attribution 4.0 International.
- [14] M López-García. Stochastic descriptors in an sir epidemic model for heterogeneous individuals in small networks. *Mathematical biosciences*, 271:42–61, 2016.
- [15] Antonella Liccardo and Annalisa Fierro. A lattice model for influenza spreading. *PLoS One*, 8(5):e63935, 2013.
- [16] Emir Baysazan, A Nihat Berker, Hasan Mandal, and Hakan Kaygusuz. Covid-19 modeling based on real geographic and population data. *arXiv preprint arXiv:2207.01585*, 2022.
- [17] Pei-Duo Yu, Chee Wei Tan, and Hung-Lin Fu. Epidemic source detection in contact tracing networks: Epidemic centrality in graphs and message-passing algorithms. *IEEE Journal of Selected Topics in Signal Processing*, 16(2):234–249, 2022.

- [18] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno. Discrete-time markov chain approach to contact-based disease spreading in complex networks. *EPL (Europhysics Letters)*, 89(3):38009, feb 2010.
- [19] Hend Alrasheed, Alhanoof Althnian, Heba Kurdi, Heila Al-Mgren, and Sulaiman Alharbi. Covid-19 spread in saudi arabia: modeling, simulation and analysis. *International Journal of Environmental Research and Public Health*, 17(21):7744, 2020.
- [20] Hamza Boutayeb, Sara Bidah, Omar Zakary, and Mostafa Rachik. A new simple epidemic discrete-time model describing the dissemination of information with optimal control strategy. *Discrete Dynamics in Nature and Society*, 2020, 2020.
- [21] Thanh Tam Nguyen, Thanh Cong Phan, Minh Hieu Nguyen, Matthias Weidlich, Hongzhi Yin, Jun Jo, and Quoc Viet Hung Nguyen. Model-agnostic and diverse explanations for streaming rumour graphs. *Knowledge-Based Systems*, 253:109438, 2022.
- [22] Ben Roberts and Dirk P Kroese. Estimating the number of st paths in a graph. *J. Graph Algorithms Appl.*, 11(1):195–214, 2007.
- [23] Paul Erdős, Alfréd Rényi, et al. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5(1):17–60, 1960.
- [24] Networkx. [https://networkx.org/documentation/stable/reference/generated/networkx.generators.random\\_graphs.erdos\\_renyi\\_graph.html](https://networkx.org/documentation/stable/reference/generated/networkx.generators.random_graphs.erdos_renyi_graph.html).
- [25] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
- [26] Networkx. [https://networkx.org/documentation/stable/reference/generated/networkx.generators.random\\_graphs.barabasi\\_albert\\_graph.html](https://networkx.org/documentation/stable/reference/generated/networkx.generators.random_graphs.barabasi_albert_graph.html).
- [27] Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’ networks. *nature*, 393(6684):440–442, 1998.
- [28] Networkx. [https://networkx.org/documentation/stable/reference/generated/networkx.generators.random\\_graphs.connected\\_watts\\_strogatz\\_graph.html#networkx.generators.random\\_graphs.connected\\_watts\\_strogatz\\_graph](https://networkx.org/documentation/stable/reference/generated/networkx.generators.random_graphs.connected_watts_strogatz_graph.html#networkx.generators.random_graphs.connected_watts_strogatz_graph).