Network-based Epidemic Model with Reinfections*

Rodrigo Tuna[190011317]

University of Bologna rodrigo.tunade@studio.unibo.it

Abstract. With the global impact of the COVID-19 pandemic, there has been an increasing interest in epidemic models. These models have proven to be crucial resources for comprehending and forecasting the spread of infectious diseases. This research introduces a unique epidemic model that combines epidemics' social and biological aspects based on a network of social relationships. In contrast to conventional models, we consider the potential for numerous infections and include dynamic aspects such as a decline in social connections due to the manifestation of symptoms. We experiment on both synthetic and real-world data and empirically investigate our model to examine how different network topologies and simulation settings affect the spread of infections. We also compare our model to the well-known SIR, SIS, and SEIR models.

Keywords: Network Science · Epidemic Models · Simulation.

1 Introduction

With the surge of the COVID-19 pandemic, there has been a renewed interest in epidemic models [10]. Epidemic models are mathematical tools used to simulate and study the spread of infectious diseases within a population. They have proved valuable in predicting the course of outbreaks, helping health organizations and governments make decisions, and creating informed public health policies that minimize the impact of an epidemic.

Many epidemic models exist in the literature, most of which are defined by differential equations [10], which track the evolution of the number of infected individuals operating at a macroscopic level. More complex models incorporate different strategies for disseminating a disease that consider the social nature of contacts. These models can be based on social networks [9] or game theory [8] and have a microscopic scope.

In this paper, we propose a new epidemic model that is based on a network of social contacts, incorporating both biological and social aspects of epidemics. Each individual can transmit the disease to a set of individuals it is connected with, taking into account the frequency of contacts. Unlike most models in the literature, we consider the possibility of individuals being infected more than once, which is a behavior exhibited by some infectious diseases [2]. Moreover, we

* Course Project for Complex Systems and Network Science, University of Bologna.

incorporate other epidemic behaviors, such as a natural decrease in social contacts due to disease symptoms, and implement different containment measures.

Experiments were performed on both real-world and synthetic data, showing an empirical study of how the simulation parameters and different network topologies affect the spread of the infection. We provide results for the impact of the aforementioned containment measures and also compare our model with known SIR, SIS, and SEIR models.

The main contributions of this paper are as follows:

- A new epidemic model that is based on a network and thus can be applied to real-world social network data.
- An empirical study of the model's behavior under varying graph topologies, simulation parameters, and initial infected nodes.
- Code implementation of the model and containment measures.

This paper starts by providing a review of epidemic models and random graph models. We then theoretically define our model, providing the governing equations. We explain the experimental setup and present the results obtained, comparing them with well-known epidemic models.

2 Related Work

In this section, we provide an overview of epidemic models and introduce two random graph models that will be used for experiments.

2.1 Epedemic Models

The study of an epidemic disease is a topic where there exists a contribution of both biological and social factors [5]. Epidemic models are mathematical models that can explain how infectious diseases progress inside a population, showing the result of an epidemic. The most common and well-studied epidemic models are compartment models. Compartment models segment the population into different compartments according to their status. The SIR model employs the use of three compartments: Susceptible, Infected, and Recovered. Infected individuals become immune to disease after recovering and thus can only be infected once. The SIS model represents diseases without permanent immunity, and infected individuals become susceptible again after recovering. More complex models, such as the SEIR [10] (susceptible-exposed-infected-recovered) or the SAIRS [13] (susceptible-asymptomatic-infected-recovered-susceptible), have surged and been used to depict more complex real-world behaviors of diseases. These models are categorized into deterministic or stochastic [3]. Deterministic models are described using a system of differential equations that depict the change of each group over time. Stochastic models are formulated using Markov Chains, representing each individual's transition probabilities. Stochastic models are more faithful since the spread of an epidemic is also stochastic, although they can become harder to analyze.

However, these models only account for the biological aspects of epidemics, as it is assumed every individual uniformly encounters others. To incorporate social factors into the model, a social network is used to disseminate the disease. Each individual has a finite set of possible contacts to whom they can spread the disease. The collection of all contacts forms a "mixing network" [9], which is the underlying structure used to study infection dynamics. Game Theory has been used to extend these models by allowing agents and decision-makers to take actions that maximize an individual or collective reward. However, game theoretical epidemic models are beyond the scope of this paper. For more information, the reader is referred to [8].

2.2 Random Graph Models

Random Graph Models are a fundamental topic of Network Science [1], which is at the intersection between Statistics and Graph Theory. The properties of graphs where nodes connect randomly are studied, and authors have tried to create graphs with properties present in real-world graphs. Random Graphs are also helpful in generating graphs for simulations, as large graphs can be easily generated.

The Erdos-Renyi model [6], often referred to as the $G(n, p)$ model, creates an entirely random graph given the number of nodes n and a probability $p \in (0, 1)$ of connecting any two nodes. In this model, edges are independent, facilitating the study of properties. The probability of generating a graph with m edges is given by $p(|E| = m) = p^m(1-p)^{\binom{n}{2}-m}$ and the average number of edges is given by $\mathbb{E}[|E|] = \binom{n}{2}p$. Graphs exhibit a binomial degree distribution; the probability of a node v having degree k is $p(deg(v) = k) = \binom{n-1}{k}p^k(1-p)^(n-1-k)$. Erdos and Renyi also prove in the original paper [6] more properties of $G(n, p)$ graphs. The degree distribution follows a Poisson distribution by fixing np as a constant and letting $n \to \infty$.

Watts and Strogatz introduce a model [15] that aims to generate highly clustered graphs with small diameters. Given the number of nodes n and an integer k, we create a k-nearest neighbor regular lattice. This lattice is highly clustered, but it does not have a small diameter. The graph diameter is decreased by rewiring every edge with probability β to another node, chosen uniformly at random. When $\beta = 1$, meaning all edges are rewired, the model degenerates into an Erdös-Rényi graph. It can be seen as the interpolation of a regular lattice with a random graph.

3 Model

We present a new epidemic disease model based on network propagation. The model falls under the category of SIRS models [4] (**S**usceptible → **I**nfected → **R**ecovered → **S**usceptible) because it considers that the immunity gained after recovering from the disease is temporary, and individuals can become susceptible again. We define the underlying structure, an undirected graph $G(V, E)$, where

each node $v_i \in V$ represents an individual and each edge $(v_i, v_j) \in E$ represents a social connection between a pair of nodes and is associated with a weight w_{ij} that indicates the probability of a social connection to transmit the disease. We define the state $S_i^t \in \{0, 1, 2\}$ which indicates the health of individual v_i at discrete time step $t \leq T$, and we distinguish between susceptible, $S_i^t = 0$, infected $S_i^t = 1$, and recovered $S_i^t = 2$. We define a subset of first infected nodes $V_0 \subseteq V$, such that $S_i^0 = 1$ if $v_i \in V_0$ and $S_i^0 = 1$ if $v_i \notin V_0$. We define the probability of recovery R_i^t of individual v_i at time step t; the probability of recovery is updated at every iteration $t > 0$ according to $R_i^{t+1} = min(R_i^t \alpha, 1)$, where α is the recovery rate. The immunity of individual v_i at time step t is I_i^t. The immunity level is updated according to $I_i^{t+1} = I_i^t \lambda$ where λ is the immunity decay rate. The model degenerates into a SIR model when considering $\lambda = 1$.

At each time step $t > 0$, infected nodes can recover with probability R_i^t and gain immunity $I_i^t = 1$. Recovered nodes can become susceptible again with probability $1 - I_i^t$. The infection is spread according to network propagation. Let $A \in \mathbb{R}^{|V|x|V|}$ be the adjacency matrix of graph $G(V, E)$,

$$A_{i,j} = \begin{cases} w_{i,j} & \text{if } (v_i, v_j) \in E, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We define the probability matrix A'^t of iteration t that represents the probability of a social contact occurring at iteration t and transmitting the disease, such as:

$$A'^t_{i,j} = \begin{cases} A_{i,j} * |R_i^t - \delta| & \text{if } S_i^t = 1, \\ A_{i,j} & \text{otherwise.} \end{cases} \tag{2}$$

Where δ is a contamination threshold. The multiplication of the edge weight by a factor when the individual that causes the interaction is infected replicates the disease's symptoms (i.e., fever), which would cause the infected individual to decrease their number of social contacts. This factor decreases when the recovery rate approaches the contamination threshold and increases when it is larger. This is because symptoms become stronger and then improve during the course of an infection.

We can then sample the transmission matrix T^t, where $T_{i,j}^t$ is 1 with probability $A'^t_{i,j}$, or 0 we probability $A'^t_{i,j}$. And define the states that can be propagated ${S'}_i^t$ where

$${S'}_i^t = \begin{cases} S_i^t & \text{if } S_i^t \neq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The state of next iteration S_i^{t+1} is

$${S'}_i^{t+1} = \begin{cases} 2 & \text{if } S_i^t = 2, \\ ((T^t{}^T + I){S'})_i & \text{otherwise.} \end{cases} \tag{4}$$

Newly infected nodes have a probability ϵ of recovery.

4 Experimental Setup

We conduct experiments to evaluate and empirically extract knowledge from our model. We assess our model using real-world data and additionally test on synthetically generated data. Additionally, we use synthetic graphs according to well-known random graph models to explore how graph properties can impact the spread of disease. Moreover, we investigate how the model behaves by varying the simulation parameters. The faithfulness of the model is evaluated by comparing it to well-known epidemic models. We implemented the model in Matlab, making code for experiments available for reproduction in `https://github.com/rodrigotuna/Network-based-Epidemic-Model`.

4.1 Datasets

We use two different real-world graph datasets from [12]. The *High Energy Physics - Phenomenology collaboration network* [11] is a collaboration network of papers submitted to arXiv under the category High Energy Physics - Phenomenology. The dataset is comprised of $12,008$ nodes and $118,521$ edges. The *LastFM Asia Social Network* [14] was obtained by collecting data from the social network LastFM. Nodes represent users from Asian countries, and edges represent mutual connections between users. The dataset is comprised of $7,624$ nodes and $27,806$ edges. We also experiment on synthetically generated graphs following the random graph models *Erdos-Renyi* [6] and Watts-Strogatz [15] with different generation parameters. We could not experiment with graphs with larger sizes because of constraints related to computational capabilities.

4.2 Parameters

We vary the simulation parameters to understand how the spread of the epidemic is affected by different recovery and immunity decay rates. These parameters also allow for the disease to be adaptable to mimic the behavior of any real-world disease. The recovery rate varies according to values 1.1, 1.3, 1.5, and 1.7, and the immunity decay rate varies between 1.0, 0.9999, 0.999, 0.99, and 0.9. Moreover, we vary the parameters of the random graph models. For the Erdos-Renyi model, the edge probability varies between 0.10, 0.25, 0.50, 0.75. In the Watts-Strogatz model, the values for k are 2, 5, 10, 25, and 50; the values for β are 0.20, 0.40, and 0.60. We vary the percentage of initially infected nodes with values 0.001, 0.01, and 0.1 and also vary the selection strategy between randomly selected nodes or the most central nodes according to eigenvalue centrality.

4.3 Containment Measures Implementation

To evaluate how the spread of an epidemic can be slowed, we chose three different containment measures highlighted in [7]: the use of masks, quarantine, and vaccination. At a given iteration of the simulation, containment measures

are imposed by artificially changing the graph or the state of the simulation. Experiments were run with intervention in iteration 5. The use of masks is implemented by decreasing all edge weights by a factor, thus reducing the chance that contact between nodes could transmit the disease. For experiments, we chose a factor equal to 0.5. Quarantine is implemented by erasing from the network all edges that are under a certain threshold, as that would mean minor contacts that would not occur given the imposition of a quarantine. The chosen threshold was 0.8. Vaccination was implemented by randomly selecting a population percentage and setting their state as immune and their immunity as 1.0. The selected percentage was 50%.

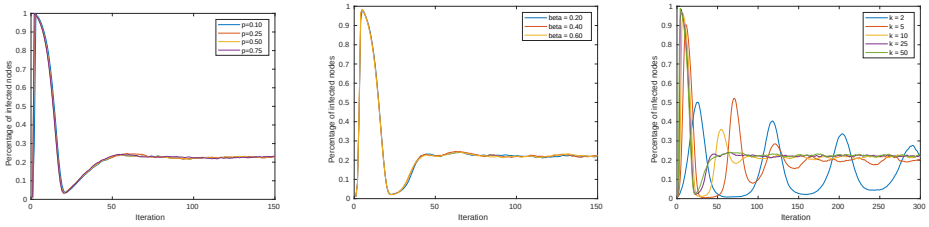
5 Results

This section provides a discussion and analysis of the different simulation runs performed. We evaluate how fast the disease is spread and what type of endemic state is reached for different scenarios. Our analysis will focus on the graph topology's impact on the disease's propagation in Section 5.1. The changes in dynamics by varying the initial set of infected nodes in Section 5.2. Section 5.3 investigates how simulation parameters can influence the spread of infection. Section 5.4 explains how different containment measures can decrease the spread of diseases. Lastly, we compare our model to the known SEIR model in Section 5.5

5.1 Topology Analysis

We analyze how the different graph properties can impact the dynamics of the infection by running simulation runs with different graph topologies. All runs were performed with the same simulation parameters, which only differed in the underlying network. Figure 1 shows that the difference in the Erdos-Renyi model's edge probability and the Watts-Strogatz model's rewiring probability have little effect on how the disease spreads across the population. All curves show a fast increase in the percentage of infected nodes, reaching almost 100%. The percentage of infections also decreases quickly and then slightly increases to stabilize in an endemic state. This leads us to conclude that the graph diameter and edge density have little effect on the spread of the disease. It is also shown that the parameter k, which relates to the average degree of the Watts-Strogatz model, significantly impacts the curve obtained. In graphs obtained with $k < 10$, there is more than one peak, and the curves oscillate around a value of infections rather than stabilizing on it. The number of peaks increases, and the size decreases when the graph's average degree decreases. This is because the average degree controls how fast the disease is spread, and in graphs with a high average degree, most nodes are infected and gain immunity after the first peak in infections. This is not the case in graphs where nodes start to recover before the disease is spread to all nodes, and thus, we have more than one infection peak.

When comparing to real-world data, a different behavior is exhibited at the peak of infections and how fast the infection is spread. Figure 2 shows that the peak in infections in real-world graphs is considerably smaller and marginally slower when compared to synthetic graphs. Despite a low average degree, the real-world graphs' degree distribution exhibits a power law, and the network still has a great capacity to spread the disease because a small percentage of nodes have a very high degree. For that reason, we only see one peak in the spread of the disease.



(a) (b) (c)

Fig. 1: Percentage of infected nodes for different types of random graph topologies with varying parameters. (a) shows simulation runs on Erdos-Renyi graphs with different edge probabilities. (b) shows simulation runs on Watts-Strogatz graphs with different rewiring probabilities. (c) shows simulation runs on Watts-Strogatz graphs with different numbers of nearest neighbor edges.

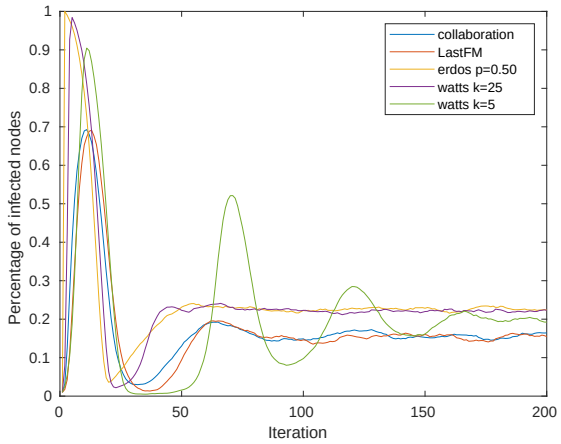
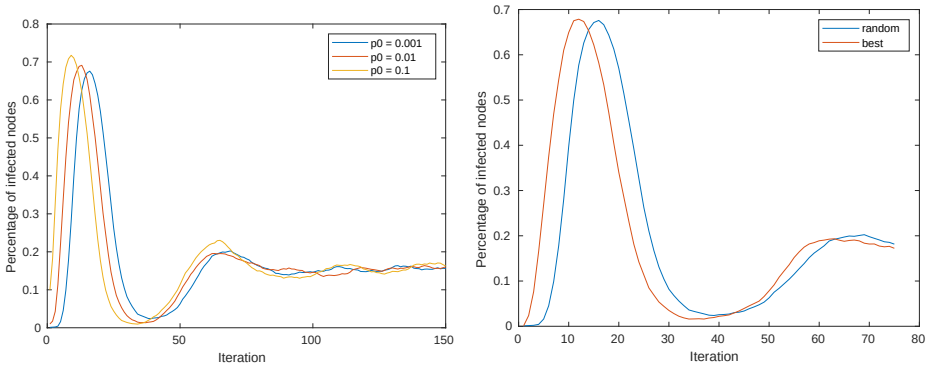


Fig. 2: Percentage of infected nodes for different graph topologies, including real-world and synthetically generated graphs.

5.2 Analysis of Initial Infected Nodes

We investigate the impact of selecting the initial set of infected nodes in disseminating a disease. We experiment with the number of initial nodes and the choice of nodes. Figure 3(a) shows different infection curves for varying percentages of randomly selected initial infected nodes. Figure 3(b) shows two different curves for percentages of infected nodes using two distinct selection strategies: randomly selected or the most central nodes. We can observe that the initial percentage of infected nodes marginally impacts the maximum number of infections and how fast the maximum is observed. The number of infections stabilizes around the same value, which is to be expected since the network and simulation parameters are the same. We also conclude that choosing the most central nodes leads to a faster peak of infections. The difference is more notable in the first five iterations of the simulation. However, there is no difference in the number of infections at the peak.



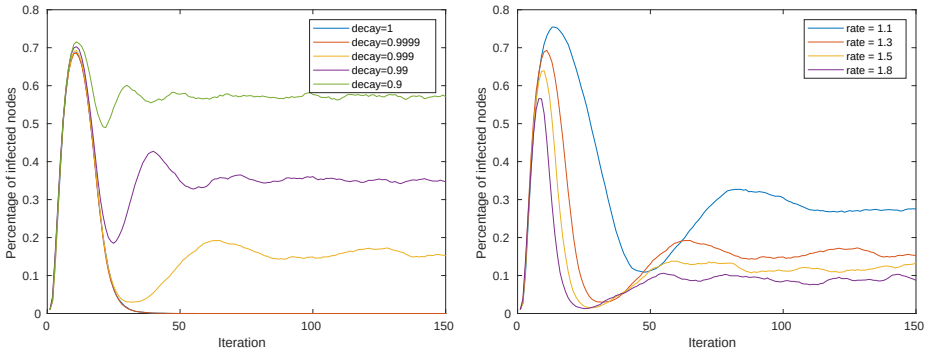
(a) (b)

Fig. 3: Percentage of infected nodes for different selections of initially infected nodes using the Lastfm dataset. (a) shows simulation runs for different initial selected percentages. (b) shows simulation runs for different selection strategies.

5.3 Parameter Analysis

We perform an analysis of how the different simulation parameters affect the propagation of the infection. We focus our analysis on the immunity decay and recovery rates, giving insights into how these parameters can be chosen to adapt our model to any real-world disease. Figure 4(a) shows the curves of infection obtained for different values of the immunity decay rate. It is observed that this parameter has no influence on the initial peak of infections and is most impactful in the stabilizing value of infections. It can also be seen that the disease cannot persist in the community for large values smaller than 1 of the

parameter and trivially when the parameter equals 1. Contrarily, Figure 4(b) shows that the different recovery rate values mainly affect the initial peak of infections. The maximum number of infections increases when the recovery rate decreases. When the disease reaches an endemic state, there is a minor variation in the value obtained at stabilization. However, it is still possible to conclude that the stabilizing value of infections grows when the recovery rate decreases. This is because the recovery rate controls the infection period of an individual.



(a) (b)

Fig. 4: Percentage of infected nodes for different selections of parameters of the model using the Collaboration dataset. (a) shows simulation runs for different immunity decay rates. (b) shows simulation runs for different recovery rates.

5.4 Efficacy of Containment Measures

This analysis focuses on how containment measures can slow down the growth of the infection or reduce the maximum number of infected nodes. Figure 5 shows different curves for distinct interventions applied at the same iteration and a curve indicating a simulation run without any intervention. All containment measures can achieve a smaller peak of infections with different degrees of success. The quarantine and vaccination significantly decreased the peak of infections, whereas the use of masks only moderately reduced them, as it is a less severe intervention. Moreover, the quarantine strategy reduced the number of contacts so the disease would die out. There were no susceptible nodes the disease could spread to before everyone recovered. Eventually, there are still susceptible nodes as the immunity decays over time, but there are no infected nodes to spread the infection. The same did not occur for other strategies where the percentage of infections in the endemic state is similar to the one reached without any intervention.

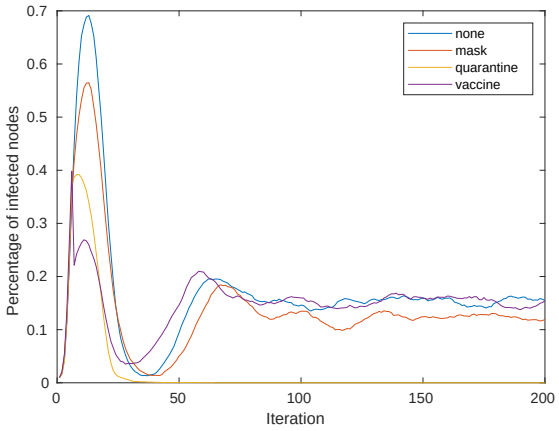


Fig. 5: Percentage of infected nodes for different containment measures using the LastFM dataset.

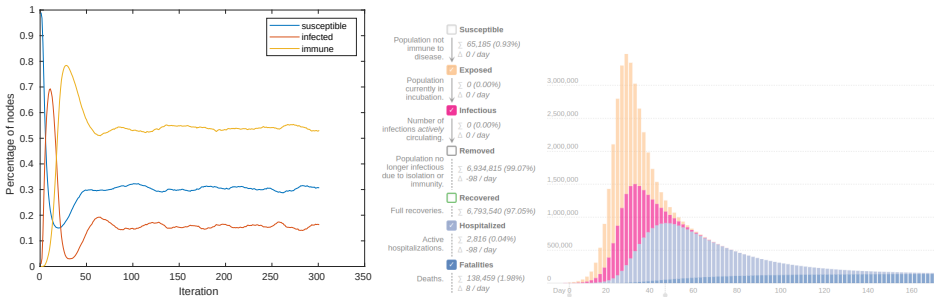
5.5 Comparison to SIS/SIR and SEIR models

We compare our work to traditional SIS/SIR [5] and the SEIR model, the last of which we consider a parametrizable implementation [1]. Typically, these models do not employ the network propagation we perform and instead rely on a system of differential equations. Both the SIR and the SEIR consider that nodes cannot be infected more than once, which leads to only one peak of infections and a subsequent decrease of infections until no node is infected. The SIS model considers that nodes do not gain immunity after infection, which results in a decrease in the susceptible population and an increase in infections until they reach an equilibrium value. Our model can be seen as an interpolation of SIS and SIR models, known as the SIRS model, as nodes gain immunity before they transit to susceptible again, resulting in one peak of infections that then decreases to stabilize in a value for the percentages of individuals in each state, reaching a stationary or oscillatory equilibrium. Furthermore, the SEIR model considers an additional state representing the incubation period. The SIR model's infected state is divided into exposed, where an individual is infected but not yet infectious, and infectious, where the individual can spread the disease to others. Another crucial difference is that unlike the models we compare our work with, our model is stochastic.

Figure 6(a) represents the evolution of a single run of our model with the different population percentages in each state. Figure 6(b) shows a SEIR model with a population size of 7,000,000,000 and $R_0 = 4.3$. When comparing the two peaks, we see a similar percentage of infected individuals at the peak, considering the total exposed and infectious. We conclude that our model overestimates the spread of infection since, although plausible, $R_0 = 4.3$ is a high value for transmission. This highlights our model's limitation since it is impossible to

[1] `https://gabgoh.github.io/COVID/index.html`

parameterize the transmission of the disease because the network dictates it. The used dataset can also be another reason for the overestimation since it is not a network of social contacts. It is also important to note that we do not assign any time period to simulation iterations, and we can vary it accordingly, allowing iterations to represent several days.



(a) (b)

Fig. 6: Comparison of our model with a SEIR model implementation. (a) shows the different percentages of individuals in each state for a simulation run. (b) shows a plot of the different states in a SEIR model with population size 7,000,000 and $R_0 = 4.3$

6 Conclusion

Epidemic models are essential for understanding an infectious disease's spread and helping decision-makers make informed policies. We produced a complete and complex model that was shown to represent the real-world dynamics of spreading diseases. The model can incorporate several aspects of social dynamics by using a contacts network for spreading the disease. Moreover, our model incorporates different behaviors present in epidemics, like reinfection and symptoms. We provide a study on the impact of the graph topology on the model, highlighting the role of the average node degree and the degree distribution in spreading an infection. We present results with different model parameters explaining the contribution of the initial choice of infected nodes and the recovery and immunity decay rates to the dynamics of the infection. We also categorize our model inside the epidemic models literature and compare it to well-established models.

Future work comprises possible extensions to the model to capture some overlooked real-world behaviors better. The model could be further improved by incorporating the other states, such as exposed, hospitalized, or dead, and considering the possibility of parameterizing the transmission rate beyond the chosen network.

References

1. Barabási, A.L., Pósfai, M.: Network science. Cambridge University Press, Cambridge (2016), http://barabasi.com/networksciencebook/
2. Bowe, B., Xie, Y., Al-Aly, Z.: Acute and postacute sequelae associated with SARS-CoV-2 reinfection. Nat Med **28**(11), 2398–2405 (Nov 2022)
3. Britton, T.: Stochastic epidemic models: A survey. Mathematical Biosciences **225**(1), 24–35 (2010). https://doi.org/https://doi.org/10.1016/j.mbs.2010.01.006, https://www.sciencedirect.com/science/article/pii/S0025556410000143
4. Cai, Y., Kang, Y., Banerjee, M., Wang, W.: A stochastic sirs epidemic model with infectious force under intervention strategies. Journal of Differential Equations **259**(12), 7463–7502 (2015). https://doi.org/https://doi.org/10.1016/j.jde.2015.08.024, https://www.sciencedirect.com/science/article/pii/S0022039615004271
5. Easley, D.A., Kleinberg, J.M.: Networks, Crowds, and Markets - Reasoning About a Highly Connected World. Cambridge University Press (2010)
6. Erdos, P., Renyi, A.: On the evolution of random graphs. Publ. Math. Inst. Hungary. Acad. Sci. **5**, 17–61 (1960)
7. Funke, M., kuang Ho, T., Tsang, A.: Containment measures during the covid pandemic: The role of non-pharmaceutical health policies. Journal of Policy Modeling **45**(1), 90–102 (2023). https://doi.org/https://doi.org/10.1016/j.jpolmod.2022.12.001, https://www.sciencedirect.com/science/article/pii/S0161893822001193
8. Huang, Y., Zhu, Q.: Game-theoretic frameworks for epidemic spreading and human decision making: A review (2021)
9. Keeling, M., Eames, K.: Networks and epidemic models. Journal of the Royal Society, Interface / the Royal Society **2**, 295–307 (06 2005). https://doi.org/10.1098/rsif.2005.0051
10. Kucharski, A.J., Russell, T.W., Diamond, C., Liu, Y., nCoV working group, C., Edmunds, J., Funk, S., Eggo, R.M.: Early dynamics of transmission and control of covid-19: a mathematical modelling study. medRxiv (2020). https://doi.org/10.1101/2020.01.31.20019901, https://www.medrxiv.org/content/early/2020/02/18/2020.01.31.20019901
11. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graph evolution: Densification and shrinking diameters (2007)
12. Leskovec, J., Krevl, A.: SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data (Jun 2014)
13. Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., Zimmer, T., Thiel, V., Janke, C., Guggemos, W., Seilmaier, M., Drosten, C., Vollmar, P., Zwirglmaier, K., Zange, S., Wölfel, R., Hoelscher, M.: Transmission of 2019-ncov infection from an asymptomatic contact in germany. New England Journal of Medicine **382**(10), 970–971 (2020). https://doi.org/10.1056/NEJMc2001468, https://doi.org/10.1056/NEJMc2001468, pMID: 32003551
14. Rozemberczki, B., Sarkar, R.: Characteristic functions on graphs: Birds of a feather, from statistical descriptors to parametric models (2020)
15. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**(6684), 440–442 (1998). https://doi.org/10.1038/30918