Evolution of an Epidemic

CSNS Project 2023/24

Maria José Carneiro

Alma Mater Studiorum - University of Bologna
maria.josecarneiro@studio.unibo.it

Abstract. The research presented in this paper explores the dynamic nature of epidemic evolution by leveraging network science. Focusing on the interplay between network topology, infection dynamics, and recovery mechanisms, the study introduces a comprehensive model incorporating elements from Erdős-Rényi random graph models and real-world face-to-face interaction datasets. Key aspects, such as edge weight evolution, methods for selecting initially infected individuals, and the impact of introducing permanent immunity, are thoroughly analyzed. By simulating the progression of an epidemic over time, the model aims to provide insights into the complex dynamics underlying infectious disease spread. Comparative analyses with traditional diffusion-based models, including SIS, SIR, and SEIR, contribute to assessing the performance and realism of the proposed approach. The outcomes highlight the need for refined weight evolution mechanisms and suggest potential improvements to enhance the model's accuracy in predicting epidemic evolution.

Keywords: Epidemic Models · Network Science · Simulation

1 Introduction

Epidemic evolution analysis explores the dynamic nature of epidemic spread through the lens of network science and how real-life phenomena can be simulated in order to predict disease contagion and, hopefully, implement measures that face it. This research delves into the intricate interplay between network topology, infection dynamics, and recovery mechanisms. The developed model incorporates elements from real-world contact networks, considering both Erdős-Rényi random graph models and actual face-to-face interaction datasets.

This paper presents a comprehensive analysis of the model, investigating critical aspects such as the evolution of edge weights, methods for selecting initially infected individuals, and the impact of introducing permanent immunity. By simulating the progression of an epidemic over time, the study aims to shed light on the complex dynamics underlying infectious disease spread. Additionally, comparisons are drawn between the developed model and traditional diffusion-based models, such as SIS, SIR, and SEIR, providing insights into the model's performance and realism.

2 Related Concepts

Epidemic infection models rely on networks to model their evolution. Due to the adaptability and versatility of using randomly generated graphs instead of fixed datasets, many graph generation models can be used to generate those graphs. One of the most well-known random graph generation models is the Erdős-Rényi network model, which builds edges between N isolated nodes with a certain predefined probability. There are many other random graph generation models like Watts-Strogatz and preferential attachment models, but I will only focus on Erdős-Rényi for this research.

In order to develop the model, it is crucial to understand how epidemics are modeled in the literature. The most popular models to simulate the evolution of epidemics are diffusion-based models, namely SIS, SIR, and SEIR. These epidemiological models can be briefly described as follows:

- **SIS** or the **Susceptible-Infectious-Susceptible** model is a network diffusion model where individuals, after being infected, remain susceptible to the disease; they can repeatedly become infected and recover with no lasting immunity. For this reason, they're appropriate for modeling recurring diseases.
- **SIR** or the **Susceptible-Infectious-Recovered** model is a network diffusion model where individuals, after being infected, can have resistance or immunity to the disease. Individuals can move through different statuses, defined by differential equations.
- **SEIR** or the **Susceptible-Exposed-Infectious-Recovered** model is a diffusion model that extends the SIR model by introducing an "Exposed" (E) compartment that represents individuals during an incubation period. Individuals move through the different states by differential equations that drive them from the susceptible state to the exposed state after exposure, then to the infectious state, and finally to the recovered state. It provides a more realistic representation of disease spread than the simpler SIR model, as it accounts for the delay between exposure to the virus and the onset of infectiousness.

These models are great tools to understand and simulate the spread of infectious diseases within populations and allow them to plan and intervene, with vaccination or social distancing, by analyzing the impact of multiple factors, like the infection rate, during an outbreak.

3 Model Definition

The model was developed considering all the project requirements with slight alterations to accommodate for more realistic elements that increase the model's complexity and fidelity to a real epidemic evolution, even within some constraints. The model definition and experiments were performed using MATLAB [2]. The following sections explain the model details according to its design.

3.1 Network Topology

The network topology was selected based on the Erdős-Rényi model and on an actual face-to-face contact network.

The Erdős-Rényi network had an initial base value of 1000 nodes and an edge formation probability of 0.01. This guaranteed the network had a reasonable size and amount of edges (around 5000), which wasn't too computationally expensive but was still representative of an epidemic evolution.

For the real-world network, my goal was to choose a dataset that closely mimicked contagion relationships since choosing larger networks (like social networks) could quickly result in overestimations, so I prioritized accuracy to the conditions instead of size. For this, I explored human contact networks, which are unipartite networks of actual contact between persons, like talking with each other, spending time together, or at least being physically close.

The network chosen was the 'Infectious' network, which describes the face-to-face behavior of people during the exhibition "INFECTIOUS: STAY AWAY"[1] in 2009 at the Science Gallery in Dublin. Nodes represent exhibition visitors, and edges represent contacts of at least 20 seconds. Since many epidemics are airborne, this dataset had great representability for the task. I pruned the dataset since it contained multiple instances of the same edges representing how long people stayed in contact. It could be relevant to model the disease spreading - longer contacts with a higher probability of transmission - but I chose not to explore that, so I removed the time column and the repeated edge instances. The final pruned dataset contained 410 nodes and 2765 edges. It is a small dataset but large compared to other contact networks I found.

Figure 1 shows the visualization of both the network models used in the experiments.

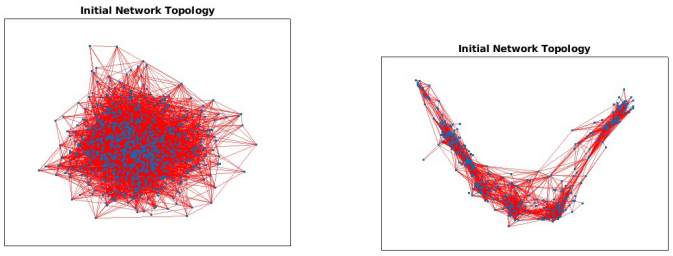

Fig. 1. Network Topology: on the left using the Erdős-Rényi model and on the right the real-world network.

3.2 Edge Weight Evolution

The edges in the model represent the probability that the disease can propagate along the edges to other nodes or, in practical terms, how likely individual A is

to transmit the disease to individual B at each time step. The edge weights are initialized with random values between 0 and 1 to represent this probability.

In real-world settings, the likelihood of infection varies with multiple factors like the time of infection, the incubation period, and containment measures, among many others. Ensuring the weights evolve during the simulation is crucial to represent the disease spread over time realistically. For this, I assumed two approaches for weight evolution in order to compare the models:

1. **Random-Weight Approach**, the baseline method, since randomly updating the weights is not representative of real-world epidemic evolution. It simply introduces some complexity and randomness to how the disease is spread.
2. **Weighted-Status Approach**, which updates edge weights according to the node infection status: edge weights are increased for infected nodes and decreased for recovered ones. In real-life terms, if an individual has a contagious disease and is infected, the likelihood of spreading the disease is much higher than for a healthy individual; vice-versa for patients who have recovered, assuming no recurring infections.
   This definition doesn't consider incubation, isolation periods, or other factors that could affect how the weights change over time. For example, the disease is dormant during incubation, but the individual is already infected. However, this simplified approach already contributes to some complexity and realism to the model.
   The magnitude of the weight changes was defined by a fixed spread parameter equal for both cases (edges increased and decreased in weights by the same ratio). To increase complexity, this parameter could vary dynamically during the simulation; in experiments, a fixed value was used for simplification purposes.

Both approaches apply the newly calculated weights to the edges of the transition matrix to affect the disease's spread.

3.3 Infection Status Node Representation

To represent infected individuals, and since the edges hold the probability of spreading the disease, a boolean representation of infected individuals was used, $I \in \{0,1\}^N, where I_i = 1$ only if the individual is infected and I_0 means that the individual is susceptible. As immunity was also modeled, further explained in section 3.5, there was a third representation of immune individuals, which is maintained in a separate vector through the simulation.

To select the initially selected individuals, I explored two approaches:

1. **Random Selection** which chose N random individuals to be the initially infected nodes that begin the spread.
2. **Centrality Selection** which considered a combined centrality of degree and closeness centrality, with equal importance to each, as a vector of the most central nodes with more edges, from where the highest N nodes were chosen as the initially infected ones.

All experiments used 5 infected individuals, as it allowed for a slower disease spread and was better to compare results.

3.4 Disease Spread Method

A transmission matrix was calculated at each time step to propagate the infection since the infected individuals are represented by a boolean vector. Given that a_{ij} are the weights of the adjacency matrix, then $T = (t_{ij})$ can be defined as:

$$T_{ij} = \begin{cases} 1 & \text{with probability } a_{ij} \\ 0 & \text{with probability } 1 - a_{ij} \end{cases} \tag{1}$$

Newly infected individuals are then calculated as $T^T I$, where I is the vector of currently infected individuals. The infection rate β is multiplied by $T^T I$ to define how quickly the disease spreads along the network.

3.5 Recoveries and Recovery Rates

Recoveries are defined by applying a fixed recovery rate δ to infected individuals. In subsequent time steps, nodes already recovered can be infected again. For this first approach, no criteria define how this disease recurrence is modeled. This could be something to improve as, realistically, an individual who has just recovered will have at least some partial immunity.

In some epidemics, vaccines can grant permanent immunity to patients who have been infected and have recovered. To model this phenomenon and compare it to the no-immunity approach, the second approach considers that recovered nodes would become permanently immune, so they could never get infected again throughout the simulation. The immune nodes are stored in a separate vector and updated at each time step with the recovered nodes. The values of those nodes are not considered when spreading the disease and are removed from the transmission matrix.

4 Model Analysis

An incremental analysis was made for each model component to evaluate each decision made to model the epidemic through time. In the following sections, comparisons between the model details are explored, as well as comparisons to SIS/SIR diffusion models.

The experiments were done in 3 different stages divided by three files MATLAB script files: `epidemic.m`, which uses the random-weight approach and doesn't define immunity; `spreading.m`, which uses the weighted-status approach and also doesn't define immunity and `immunity.m`, that uses the weighted-status approach and defines permanent immunity.

The stop criterion defined for the simulation is a fixed value. The time steps represent weeks instead of days since individuals have many contacts between each time step.

Network visualization graphs are shown at each time step in small intervals to monitor the evolution of the epidemic throughout the simulation so that disease spread is somehow animated and it's easy to see its evolution visually. At the end of the simulation, a plot of the epidemic curve is shown to visualize how the number of individuals by each disease status varies over the time steps.

4.1 Analysis of Weight Evolution Approach

Experimental Setup This analysis was done on the `epidemic.m` model that uses the random weight approach and on the `spreading.m` model that uses the weighted-status approach. It was performed using the Erdős-Rényi network with 1000 nodes and a 0.01 edge probability for both cases. The same parameters were used on both the models evaluated for the random-weight and weighted-status approach: recovery rate is 0.05, infection rate is 0.3, and weight spread parameter is 0.01.

Comparison As observed in Figure 2, comparatively, both models behave similarly in the short term (10 iterations) or in the long term (100 iterations). This shows that the chosen spread mechanism might not represent the contagion dynamics affecting the disease spread since it has the same behavior as a completely random model. Increasing/decreasing by a fixed spread parameter creates long-term oscillatory behaviors, meaning that evolution is stuck in a feedback loop of increasing and decreasing with no meaningful dynamics, making it no better than using completely random weights at each time step. In fact, although recovering is implemented, since there is no immunity, this becomes similar to an SI model where individuals stay infected and infectious throughout their lives and remain in contact with the susceptible population.

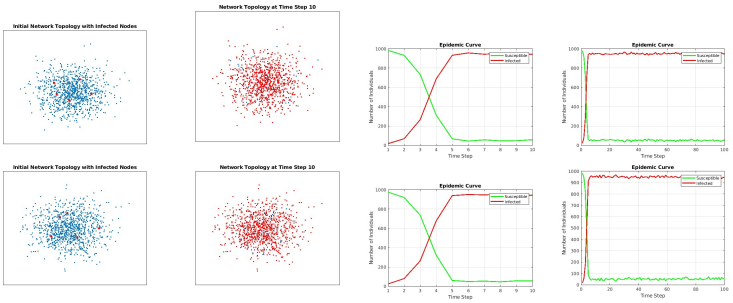

Fig. 2. Analysis of Weight Evolution: above is the Random Weight approach and bellow the Weight-Status approach, for the Erdős-Rényi network

4.2 Analysis of Infected Selection Approach

Experimental Setup This analysis was done on the `spreading.m` model that uses the weighted-status approach and doesn't define immunity. It was performed using the real-world network. The same parameters were used on both the models evaluated for the random and the centrality selection: recovery rate is 0.05, infection rate is 0.3, and weight spread parameter is 0.01.

Comparison As observed by the epidemic curve in Figure 3, the Centrality Selection makes the epidemic evolve faster since both the number of infected nodes increases and the number of susceptible nodes decreases at an earlier time step than with the Random Selection. This makes sense considering more central nodes with more edges or, in practical terms, have more contacts after they are infected, so they spread the disease more quickly.

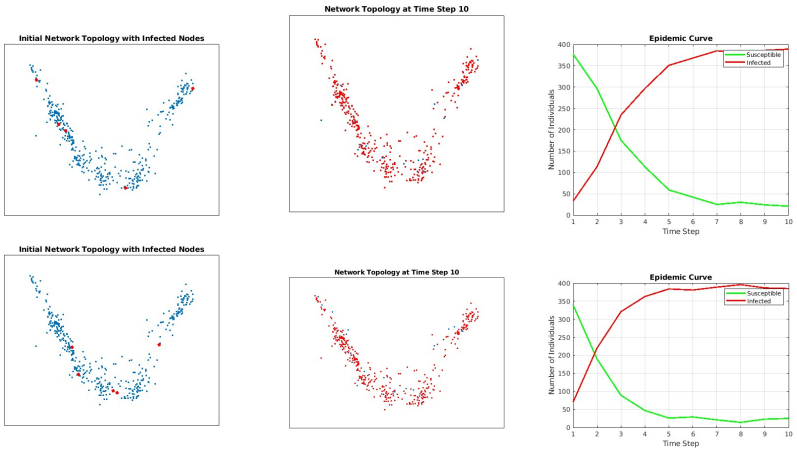

Fig. 3. Analysis of Infected Selection Approach: above is the Random Selection approach and bellow the Centrality Selection approach, for the real-world network

4.3 Analysis of Permanent Immunity

Experimental Setup This analysis was done on the `spreading.m` model that does not represent immunity and on the `immunity.m` that does describe permanent immunity. It was performed using the Erdős-Rényi network with 1000 nodes and a 0.01 edge probability for both cases. The same parameters were used on both the models evaluated for the random-weight and weighted-status approach: recovery rate is 0.05, infection rate is 0.3, and weight spread parameter is 0.01.

Comparison As observable in Figure 4, incorporating an immunity mechanism in the model creates a more realistic scenario than a permanent oscillating infection, as seen in the other models. Even though there is an early spike in infected individuals at around week 5, immune individuals steadily grow over time until, eventually, every person in the population has been infected and become immune. This is proven by the long-term range (100 iterations/weeks); the permanent immunity model evolves to a close to the entirely immune population. In the first case, recovered nodes eventually get infected again, causing the oscillations and not the equilibrium obtained in the permanent immunity model. In fact, the model follows the same distribution as a SIR model, where an epidemic eventually dies out due to insufficient susceptible individuals to sustain the disease. Infected individuals who are added later will not start another epidemic due to the lifelong immunity of the existing population.

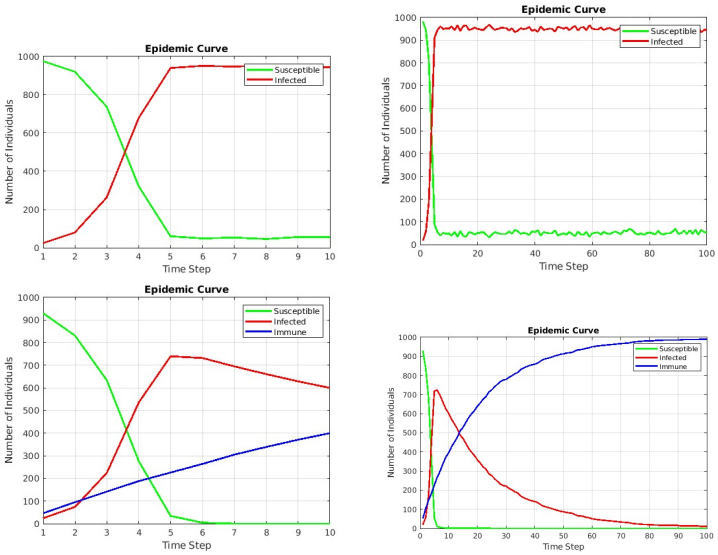

Fig. 4. Analysis of Permanent Immunity: above is the model without permanent immunity bellow is the model with permanent Immunity, for the Erdős-Rényi network

4.4 Analysis of Network Topology

Experimental Setup This analysis was done on the `immunity.m` that describes permanent immunity with Random Initial Selection and Weight-Status Evolution. It was performed using the Erdős-Rényi network with 1000 nodes, a 0.01 edge probability, and the real-world face-to-face contact network. The same parameters were used on both the models evaluated for the random-weight and

weighted-status approach: recovery rate is 0.05, infection rate is 0.3, and weight spread parameter is 0.01.

Comparison As observable in Figure 5, the evolution of the number of infected individuals is similar between both networks, although less accentuated in the real network case. This could be due to the topology of the actual network, as it has a V shape, and nodes are farther away from each other, making fewer potentially contagious contacts. The size of the networks is also different: the Erdős-Rényi network has 1000 nodes, and the actual network only has 410, which could also be the reason for the differences between the evolution.

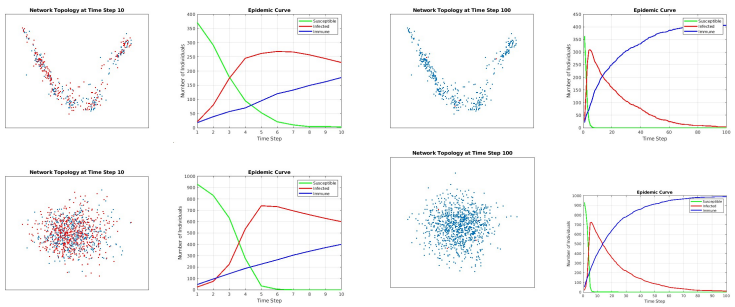

Fig. 5. Analysis of Network Topology: above is the Real-World Network and bellow the Erdős-Rényi network

5 Outcomes and Improvements

The primary concern about the developed model is the weight evolution, as it does not cause any difference in the disease-spreading mechanism and should be more representative of the infected, recovered, and immune status. Introducing some mechanism to handle partial immunity and model recurring diseases through the weight update could make the network model more realistic and predict the evolution of an epidemic more accurately. Many more randomness sources could be implemented, making the system increasingly complex and hard to predict, but certainly more in line with a real disease progression. Although simplistic, the model showed real-world phenomena like improved immunity through recovery and the effects of large contact networks in spreading disease, for example. The analysis of results, even the ones that weren't as satisfactory, was relevant to understanding and interpreting unknown and unexpected behaviors so even though some parameters could be improved, the outcomes of the research implementation were overall positive.

References

1. Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J.-F., and Van den Broeck, W. What's in a crowd? analysis of face-to-face behavioral networks. *Journal of Theoretical Biology 271*, 1 (Feb. 2011), 166–180.
2. MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2010.