Advances in Differential Equations and Control Processes 2026, 33(1), 3966.
https://doi.org/10.59400/adecp3966

Article

SI/SIS/SIR models for malware propagation in P2P networks: Numerical analysis and perspectives for fractional-order extensions

Dušan Džamić* , **Aleksa Marković**

Faculty of Organizational Sciences, University of Belgrade, Jove Ilića 154, 11010 Belgrade, Serbia

*** Corresponding author:** Dušan Džamić, dusan.dzamic@fon.bg.ac.rs

CITATION

Džamić D, Marković A. SI/SIS/SIR models for malware propagation in P2P networks: Numerical analysis and perspectives for fractional-order extensions. Advances in Differential Equations and Control Processes. 2026; 33(1): 3966. https://doi.org/10.59400/adecp3966

ARTICLE INFO

Received: 27 January 2026
Revised: 10 March 2026
Accepted: 13 March 2026
Available online: 20 March 2026

Abstract: In this paper, we examined the suitability of epidemic models on networks to assess their potential application for detecting malware propagation patterns in peer-to-peer (P2P) computer networks. We analyzed how the Susceptible-Infected (SI), Susceptible-Infected-Susceptible (SIS), and Susceptible-Infected-Recovered (SIR) models, which were originally developed for biological viruses, can be applied to digital viruses. Using the Gnutella network dataset as a representative topology of P2P networks, we simulated infection scenarios to evaluate how scale-free network properties and the presence of high-degree nodes acting as super-spreaders influence the propagation speed and network saturation. The obtained results show that the examined models can be used and provide valuable insight into epidemic dynamics. However, the existing models are not perfect, and the introduction of additional states, such as L for latency and Q for quarantine, is proposed, since these are relevant for digital devices and digital viruses. More precisely, the absence of latent (L) and quarantine (Q) components leads to an overestimation of infection speed and an inability to model strategic isolation. Accordingly, this study provides empirical evidence that standard biological models are not sufficient for accurate predictions in the field of cybersecurity in P2P environments, and that future modeling efforts should move from basic compartmental models toward more advanced frameworks, such as SEIR and SIQR, to realistically capture malware activation delays and the impact of active defense strategies.

Keywords: malware propagation; peer-to-peer networks; SI/SIS/SIR models; gnutella; network topology; super-spreaders

1. Introduction

The development of digital systems has often been influenced by biological systems. This connection has also encouraged comparisons between malicious software and organic pathogens. Since the 1980s, especially after the appearance of the Brain virus for personal computers, researchers have explored different epidemiological approaches in order to understand how computer viruses and other forms of malicious code spread [1,2]. Over time, models such as Susceptible-Infected (SI), Susceptible-Infected-Susceptible (SIS), and Susceptible-Infected-Recovered (SIR) became standard tools for estimating the scale of an outbreak and the speed of its spread in cyberspace [3–5].

While these models offer a solid mathematical basis for studying epidemics in homogeneous populations, their direct use in modern network architectures, especially

peer-to-peer (P2P) systems, is not straightforward. P2P networks are decentralized and operate without central control. They also have complex structural properties, including power-law degree distributions and highly connected nodes, often referred to as hubs [6, 7]. Such nodes can act as super-spreaders and significantly influence the propagation process. Because of that, the usual assumption of homogeneous mixing is no longer appropriate. Pastor-Satorras and Vespignani [8] pointed out that in scale-free networks the epidemic threshold may disappear, which means that infections can persist even when transmission rates are very low.

Although network epidemiology has been widely studied, there is still a clear lack of validation of classical compartmental models on real P2P topologies under different defense conditions. Many existing studies are based on random graph approximations, such as Erdős-Rényi networks, but these do not reflect the actual structural weaknesses of real file-sharing systems such as Gnutella [9]. In addition, the standard SI, SIS, and SIR models usually do not include the latency of modern malware, which may stay inactive for some time in order to avoid detection. They also do not represent the active role of quarantine measures used by antivirus programs and firewalls.

In this paper, we address these limitations through a numerical analysis of malware propagation on the Gnutella network topology. We simulate the behavior of the SI, SIS, and SIR models in order to examine how suitable they are in a digital setting. A central part of the study is the comparison of different infection scenarios. In particular, we analyze the difference between outbreaks that start from randomly selected nodes and those initiated by super-spreaders.

The main contributions of this study can be summarized as follows:

1. We carry out a numerical analysis of the SI, SIS, and SIR models on a real unstructured P2P network. This allows us to examine how differences in network topology affect the speed of malware propagation.
2. We study the super-spreader effect by comparing outbreaks that begin at highly connected nodes with those that start at randomly chosen nodes. The results show that infections initiated at high-degree nodes spread through the network much faster.
3. Based on the obtained results, we point out that basic compartmental models are not sufficient to describe all relevant aspects of malware behavior. In particular, the findings support the inclusion of additional states such as Latent (L) and Quarantine (Q), as in SEIR and SIQR models, in order to obtain more realistic predictions in cybersecurity applications.

The rest of the paper is organized as follows. Section 2 presents the mathematical background of the epidemiological models and introduces the Gnutella network dataset. Section 3 describes the simulation procedure, the selected parameter values, and the numerical results. Section 4 discusses the limitations of the considered models and their implications for control strategies. Section 5 gives the concluding remarks and outlines possible directions for future research.

2. Mathematical preliminaries and network topology

In this section, we formulate the compartmental epidemiological models used to describe the propagation of malware and define the topological properties of the peer-to-peer network under study. We consider a network represented by a graph $G = (V, E)$, where V is the set of nodes (computers) with $|V| = N$, and E is the set of edges (connections) representing communication channels through which malware can spread.

2.1. Epidemic modeling

In order to model the spread of epidemics effectively, we must first refer to two fundamental hypotheses that form the basis of the analytical and numerical framework for epidemic modeling, namely the hypothesis of "states of individuals during an epidemic" and the homogeneous mixing hypothesis.

Models based on the hypothesis of "states of individuals during an epidemic" place each individual into one of several categories according to their condition. The simplest model defines three states in which an individual may be found during an epidemic:

- Susceptible (S),
- Infected (I), and
- Removed (immune/deceased) (R).

The susceptible state defines healthy individuals who, up to a given moment, have not been in contact with the pathogen. The infected state includes individuals who have been in contact with the virus and are therefore capable of infecting other previously uninfected individuals. Individuals who were infected at some point in the past but have recovered are placed in the removed category, with the emphasis that they are no longer infectious and cannot be infected again. These three categories are not always sufficiently specific for modeling some other pathogens. For the purpose of modeling more complex diseases, additional states may also be used, such as the latent state (this state includes individuals who have been exposed to the pathogen but have not developed the disease) and the immune state (this state describes individuals who cannot be infected by the given virus) [7]. For the sake of modeling accuracy, as well as for consistency with the real world, individuals may change states during the course of an epidemic. If we take a virus as an example, then during some unspecified period of time, an individual who is currently not infected (that is, is in the susceptible state) may come into contact with another infected individual and move to the infected state. Likewise, through appropriate actions and after a certain period of time, that individual may move to the recovered state and very likely develop certain characteristics that increase the ability of the immune system to defend against a similar strain of the virus in the future.

The homogeneous mixing hypothesis assumes that each individual has approximately the same chance of coming into contact with another infected individual. This hypothesis also removes the need for a precise definition of the contact network on which the virus spreads. Instead of an explicitly defined contact network, the basic assumption is that each individual may infect any other individual.

This hypothesis is also known as the complete mixing hypothesis [7].

We analyze the dynamics of three fundamental models: SI, SIS, and SIR. Let $S(t)$, $I(t)$, and $R(t)$ denote the fraction of susceptible, infected, and recovered nodes at time t, respectively, such that $S(t) + I(t) + R(t) = 1$. The transmission rate is denoted by β, and the recovery rate by γ.

2.1.1. The SI model

The Susceptible-Infected (SI) model describes pathogens for which there is no immunity or recovery; once a node is infected, it remains infected forever. This is similar to malware that permanently compromises a system and cannot be removed. Its dynamics are described by the following differential Equation (1):

$$\frac{dI}{dt} = \beta I(t)S(t). \tag{1}$$

Because $S(t) = 1 - I(t)$, the solution has the form of a logistic growth curve:

$$I(t) = \frac{I_0 e^{\beta t}}{1 - I_0 + I_0 e^{\beta t}}, \tag{2}$$

where I_0 is the initial fraction of infected nodes. In a finite network, the infection eventually permeates the entire population, i.e., $\lim_{t\to\infty} I(t) = 1$.

Following the framework established by Barabási and Albert [7], the spreading rate in a network is determined not just by β, but by the network's moment statistics. For a homogeneous random network, the characteristic time scale is $\tau = \frac{1}{\beta\langle k\rangle}$. However, Gnutella exhibits a heterogeneous degree distribution. In such scale-free topologies, the infection rate is dominated by the most connected hubs, leading to a drastically reduced characteristic time $\tau \approx \frac{\langle k\rangle}{\beta\langle k^2\rangle}$. This theoretical insight explains the rapid saturation observed in our super-spreader simulations (Section 4), as the second moment $\langle k^2\rangle$ diverges in power-law networks.

2.1.2. The SIS model

The Susceptible-Infected-Susceptible (SIS) model accounts for scenarios where nodes can recover from infection but do not acquire permanent immunity (e.g., malware is removed, but the system remains vulnerable to reinfection). The system is described by:

$$\frac{dI}{dt} = \beta I(t)S(t) - \gamma I(t). \tag{3}$$

This model exhibits a threshold behavior. If the basic reproduction number $R_0 = \beta/\gamma > 1$, the infection survives and reaches an endemic steady state. If $R_0 \leq 1$, the infection dies out exponentially.

A critical distinction in network epidemiology is the epidemic threshold λ_c, below which the virus dies out. While in homogeneous networks $\lambda_c = \frac{1}{\langle k\rangle}$, Barabási and Albert [7], and Pastor-Satorras and Vespignani [8] proved that for scale-free networks with degree exponent $2 < \alpha \leq 3$ (typical for P2P networks), the threshold vanishes:

$$\lambda_c = \frac{\langle k\rangle}{\langle k^2\rangle} \to 0 \quad \text{as} \quad N \to \infty. \tag{4}$$

This implies that in the Gnutella network, the SIS model predicts a persistent endemic state for virtually any non-zero transmission rate β, as confirmed by our numerical results, where infections persist even at low transmission weights.

2.1.3. The SIR model

The Susceptible-Infected-Recovered (SIR) model introduces a removed state R, representing nodes that are immune, patched, or disconnected from the network. The system of coupled differential equations is:

$$\frac{dS}{dt} = -\beta S(t)I(t) \tag{5}$$

$$\frac{dI}{dt} = \beta S(t)I(t) - \gamma I(t) \tag{6}$$

$$\frac{dR}{dt} = \gamma I(t). \tag{7}$$

In the SIR model, the epidemic eventually terminates as the pool of susceptible nodes is depleted, i.e., $\lim_{t\to\infty} I(t) = 0$.

In the SIR model applied to complex networks, the timing of the outbreak is heavily dependent on the connectivity of the initial node. As described in the previous studies [7, 10], hubs are statistically likely to be infected early in the process. Once infected, a hub with degree k_{hub} broadcasts the virus to k_{hub} neighbors simultaneously. This network property creates a kind of hierarchical cascade. The virus first spreads from the super-spreader to intermediate hubs, and then from those hubs to peripheral nodes. As a result, the time needed to reach peak infection is much shorter than predicted by the mean-field approach.

2.2. Peer-to-peer networks

Peer-to-peer networks represent a decentralized model composed of interconnected nodes, that is, computers that share resources. In this model, there is no need for a centralized server or any form of centralized administration. The idea of peer-to-peer networks dates back to the 1960s, when a group of experts developed ARPANET with that concept in mind. However, peer-to-peer networks gained much greater importance much later, at the end of the 1990s, with the emergence of file-sharing systems such as Napster. During the same period, the decentralized platform Freenet was also developed. This platform was created with the intention of building a system resistant to censorship, with anonymous communication as its main goal.

In peer-to-peer networks, each node acts simultaneously as both a server and a client, which makes the network generally self-sustaining. This is the main difference compared to the client-server model, where one or more nodes in the network are identified as server nodes, that is, nodes that provide resources, while the remaining nodes are identified as clients, meaning that they act as consumers in that network.

Peer-to-peer networks offer several advantages over the traditional client-server architecture, such as greater fault tolerance, easier scalability, and potentially lower infrastructure maintenance costs. Since a centralized server does not exist in this architecture, the failure of one node often does not have consequences for the entire

network. Moreover, peer-to-peer networks can dynamically expand or contract depending on the needs of users, that is, nodes, which leads to a high level of utilization of available resources.

However, the same properties that represent advantages of peer-to-peer networks over traditional architecture also bring certain challenges, especially in the field of security [11, 12]. The lack of centralized control can allow malicious software, or malware, to spread more easily through the network.

2.3. The Gnutella dataset

To test the proposed models in a setting that is closer to a real digital environment, we use the Gnutella peer-to-peer network dataset. The open architecture, large scale, and self-organizing structure of the Gnutella network make it a particularly suitable P2P system for this type of analysis. Like many other peer-to-peer applications, Gnutella builds, at the application level, a virtual network with its own routing mechanisms. The topology of the virtual network and its routing rules have a significant influence on important system properties (i.e., performance, reliability, scalability) and potential for malware propagation.

Gnutella graph has $N = 10{,}876$ nodes and $|E| = 39{,}994$ edges (**Figure 1**). The graph is weakly connected, while the largest strongly connected component contains 4,317 nodes (less than 40% of all nodes).

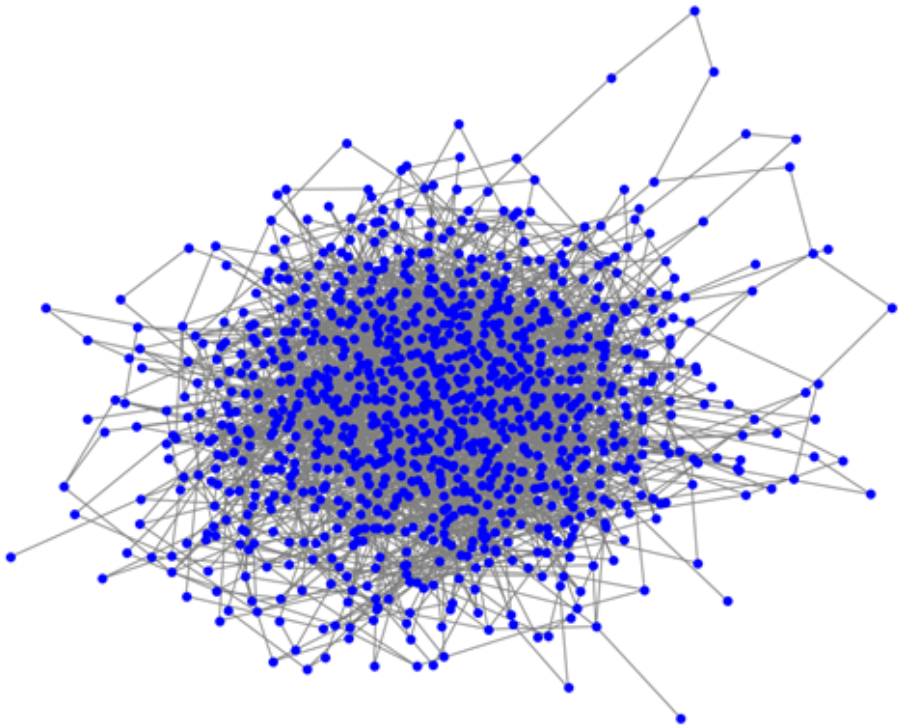

Figure 1. Visual representation of the Gnutella network.

These topological factors reinforce the network as small world type which can also be relevant in epidemic investigation. Network diameter (longest shortest path) is 9. The 90% effective diameter is only 5.4, indicating that infections can travel relatively fast through the network. Likewise, local connectivity is limited. The average

clustering coefficient is 0.0062, while the fraction of closed triangles is 0.0018. In total, the network has 934 triangles. This indicates that the local redundancy is far lower than common social networks, but still contributes to efficient propagation.

Empirical studies of Gnutella have also shown that its degree distribution follows a power-law pattern, with exponent values in the interval $\alpha \in [2.0, 2.3]$ [9, 13]. This places the network in the class of anomalous networks, defined by Cohen and Havlin [14] as networks for which $2 < \alpha \leq 3$. In that regime, the second moment of the degree distribution diverges, that is, $\langle k^2 \rangle \to \infty$ as the network size increases. As discussed in Section 2.1, this property is enough, from a theoretical point of view, to eliminate the epidemic threshold. As a result, the network becomes highly vulnerable to persistent infections, even when the transmission rate is low.

Identification of super-spreaders

The malware transmission rate is not exclusively determined by the intrinsic virulence of the software code itself but by the network topology, including the presence of high-degree nodes referred to as "super-spreaders". In this paper, the super-spreader is defined by means of degree centrality (number of active connections), which indicates the point in the top 5% of the network. And there were 543 such nodes identified in our analysis of the Gnutella dataset serving as central nodes for both information and malware propagation. The node with ID 3109 emerged as the most significant super-spreader, with the highest degree in the network (103 direct connections). This is much greater than the average for the network; therefore, node 3109 (**Figure 2**) is an ideal vector for the quick arrival of a large-scale epidemic. In our simulation, we specifically work with node 3109 to simulate a "worst-case scenario" attack vector, compared with a random infection scenario.

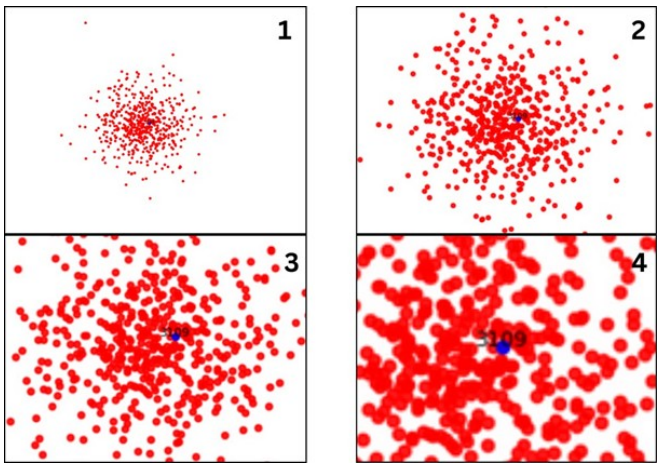

Figure 2. Node (ID 3109) and its direct connections.

3. Experimental results

This section presents the comparative results of the SI, SIS, and SIR models. We investigate how changes in the network topology, as well as epidemiological factors, affect the final epidemic size, maximal infection, and spread velocity.

3.1. Simulation framework

To evaluate the propagation dynamics of malware, we utilized the EoN (Epidemics on Networks) Python library [15]. Simulations were performed on a computer with an AMD Ryzen 5 5600 processor and 32 GB of RAM.

We specify two different scenarios of simulation given that the vector of infection has an initial presence:

1. **Scenario A (super-spreader):** The infection is initiated by highest-degree node 3109 ($k = 103$).
2. **Scenario B (random node):** The infection is initiated by low-degree node 1.

For each scenario, we varied the transmission rate (β) and recovery rate (γ) to model different categories of malware and defense capabilities. The time horizon for all simulations was set to $t_{max} = 100$ time units.

3.2. Analysis of the SI model

The SI model assumes no recovery, representing a worst-case scenario. As shown in **Tables 1** and **2**, the entire network eventually becomes infected ($N = 10,876$). However, the speed of saturation differs.

Table 1. SI model results: Time to full saturation for infections initiated by super-spreader node 3109.

Transmission rate	Final epidemic size	Peak epidemic size	Time to peak
0.2	10,876	10,876	45.33
0.4	10,876	10,876	22.86
0.6	10,876	10,876	14.71
0.8	10,876	10,876	11.74
1.0	10,876	10,876	8.99

Table 2. SI model results: Time to full saturation for infections initiated by random node 1.

Transmission rate	Final epidemic size	Peak epidemic size	Time to peak
0.2	10,876	10,876	44.25
0.4	10,876	10,876	24.16
0.6	10,876	10,876	14.76
0.8	10,876	10,876	12.00
1.0	10,876	10,876	9.26

Figures 3 and **4** demonstrate that for high transmission rates ($\beta > 0.4$), the topology has little effect on the final outcome, as the network is overwhelmed almost instantly.

3.3. Analysis of the SIS model

The SIS model represents scenarios where malware can reinfect nodes after removal (e.g., persistent botnets). We executed simulations with $\beta \in [0.2, 1.0]$ and $\gamma \in [0.1, 0.9]$.

Tables 3 and **4** summarize the average results for the super-spreader and random

node scenarios, respectively.

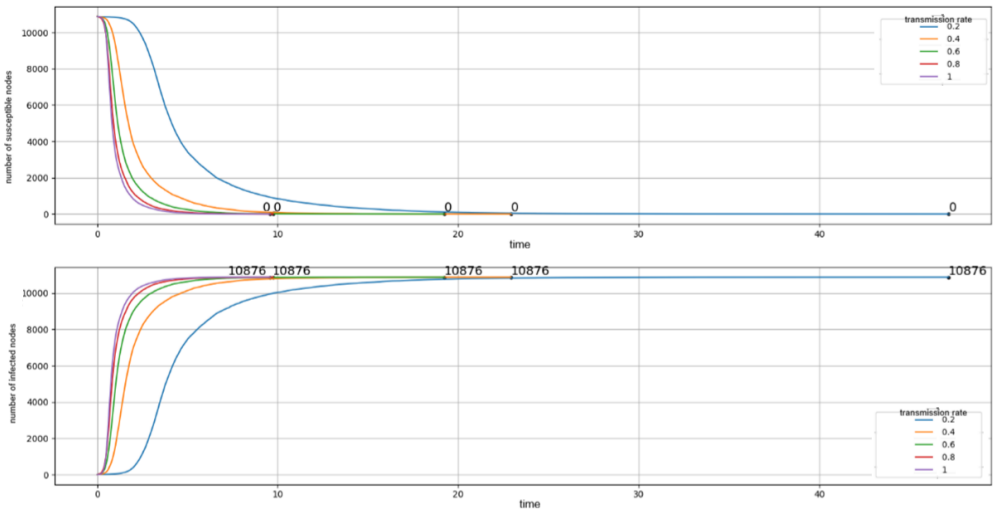

Figure 3. SI time series with initial infection at super-spreader node 3109.

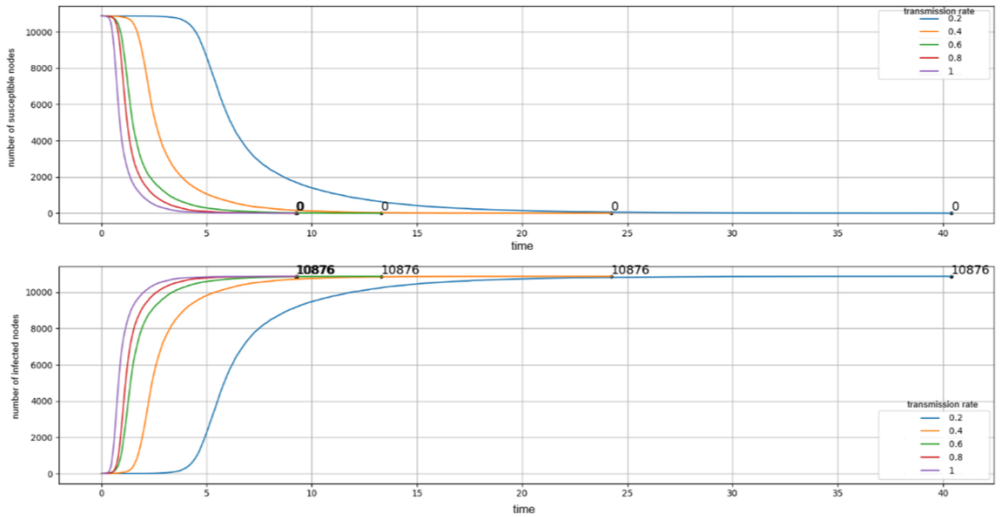

Figure 4. SI time series with initial infection at random node 1.

Table 3. SIS model results: Complete parameter sweep for infections initiated by super-spreader node 3109.

Rate β	Parameters γ	Final size	Peak size	Time to peak	Avg. infection duration $(1/\gamma)$	Reprod. number (R_0)
0.2	0.1	9234.0	9346.2	61.06	10.00	2.00
0.2	0.3	7309.6	7466.6	52.48	3.33	0.67
0.2	0.5	5987.0	6162.0	45.65	2.00	0.40
0.2	0.7	4972.0	5148.2	44.12	1.43	0.29
0.2	0.9	2514.4	2607.2	32.90	1.11	0.22
0.4	0.1	9957.8	10,065.0	69.35	10.00	4.00
0.4	0.3	8659.4	8794.4	74.42	3.33	1.33
0.4	0.5	7684.4	7859.8	48.60	2.00	0.80
0.4	0.7	6942.4	7088.6	64.33	1.43	0.57
0.4	0.9	5017.2	5174.0	52.41	1.11	0.44
0.6	0.1	10,234.2	10,323.8	51.60	10.00	6.00
0.6	0.3	9274.8	9380.6	64.97	3.33	2.00

Table 3. *Cont.*

Rate β	Parameters γ	Final size	Peak size	Time to peak	Avg. infection duration $(1/\gamma)$	Reprod. number (R_0)
0.6	0.5	8504.8	8620.2	34.55	2.00	1.20
0.6	0.7	7829.8	7992.8	43.87	1.43	0.86
0.6	0.9	7265.0	7465.8	35.99	1.11	0.67
0.8	0.1	10,402.2	10,462.0	48.85	10.00	8.00
0.8	0.3	9579.6	9701.0	55.13	3.33	2.67
0.8	0.5	8950.0	9101.0	54.49	2.00	1.60
0.8	0.7	8420.4	8554.8	46.14	1.43	1.14
0.8	0.9	7925.8	8087.0	36.55	1.11	0.89
1.0	0.1	10,488.0	10,551.2	46.46	10.00	10.00
1.0	0.3	9833.6	9915.0	33.30	3.33	3.33
1.0	0.5	9257.0	9390.4	32.42	2.00	2.00
1.0	0.7	8788.6	8923.8	68.40	1.43	1.43
1.0	0.9	8325.2	8490.8	48.66	1.11	1.11

Table 4. SIS model results: Complete parameter sweep for infections initiated by random node 1.

Rate β	Parameters γ	Final size	Peak size	Time to peak	Avg. infection duration $(1/\gamma)$	Reprod. number (R_0)
0.2	0.1	9265.0	9364.6	82.60	10.00	2.00
0.2	0.3	5823.8	5962.8	34.55	3.33	0.67
0.2	0.5	4803.4	4958.4	49.28	2.00	0.40
0.2	0.7	4015.6	4127.6	41.37	1.43	0.29
0.2	0.9	4147.0	4337.8	80.49	1.11	0.22
0.4	0.1	9966.4	10,062.2	44.86	10.00	4.00
0.4	0.3	5182.4	5278.2	32.22	3.33	1.33
0.4	0.5	7734.0	7871.4	60.08	2.00	0.80
0.4	0.7	6936.4	7107.2	50.12	1.43	0.57
0.4	0.9	2536.8	2582.8	27.61	1.11	0.44
0.6	0.1	10,245.4	10,325.8	46.59	10.00	6.00
0.6	0.3	9245.6	9372.0	66.62	3.33	2.00
0.6	0.5	6804.8	6909.4	40.79	2.00	1.20
0.6	0.7	7858.4	8000.4	57.42	1.43	0.86
0.6	0.9	7313.6	7459.6	45.62	1.11	0.67
0.8	0.1	10,393.0	10,465.2	76.26	10.00	8.00
0.8	0.3	9569.4	9703.4	46.11	3.33	2.67
0.8	0.5	7162.0	7260.0	44.83	2.00	1.60
0.8	0.7	8444.0	8540.6	58.64	1.43	1.14
0.8	0.9	6310.0	6461.4	45.02	1.11	0.89
1.0	0.1	10,470.0	10,551.4	52.88	10.00	10.00
1.0	0.3	9810.2	9923.4	46.91	3.33	3.33
1.0	0.5	9242.0	9385.6	37.10	2.00	2.00
1.0	0.7	8777.6	8920.8	36.84	1.43	1.43
1.0	0.9	8353.6	8503.4	81.04	1.11	1.11

The results indicate a strong positive correlation between transmission rate and epidemic size. Crucially, when γ increases from 0.1 to 0.9 (at $\beta = 0.2$), the final epidemic size drops by approximately 72%, highlighting the efficacy of rapid recovery mechanisms.

Figures 5 and **6** illustrate the time series of infection. In the super-spreader scenario, the infection reaches its peak significantly faster due to the hub's high connectivity.

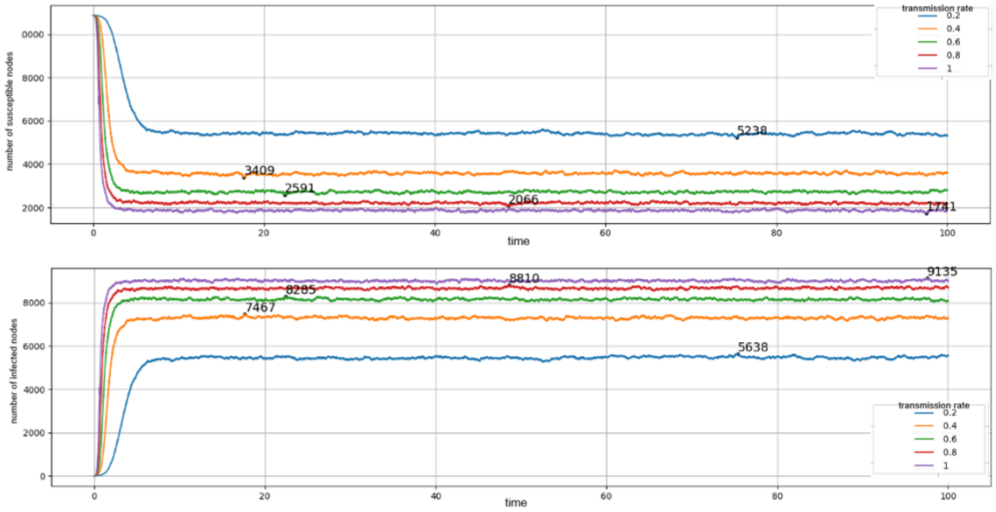

Figure 5. SIS time series with initial infection at super-spreader node 3109.

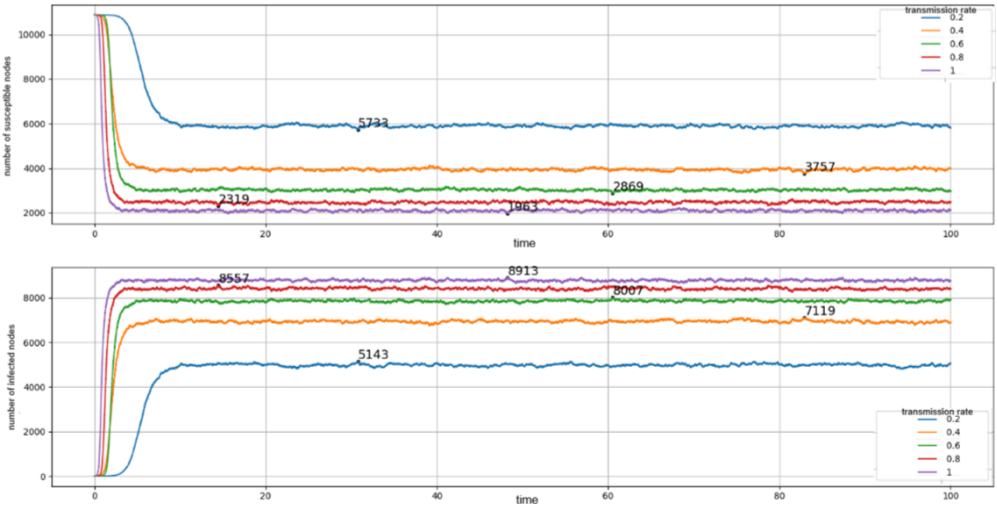

Figure 6. SIS time series with initial infection at random node 1.

3.4. Analysis of the SIR model

The SIR model introduces permanent immunity. Unlike SI and SIS, the SIR model shows that malware outbreaks can be self-limiting. **Tables 5** and **6** reveal that under many parameter combinations, the final epidemic size is negligible (0), suggesting the infection dies out before becoming endemic.

A critical observation is the "super-spreader volatility." At intermediate recovery rates (e.g., $\beta = 0.4, \gamma = 0.7$), initiating the infection at node 3109 results in a peak size of 3687 nodes, whereas initiating at node 1 results in only 2215 nodes. This confirms that targeted attacks on hubs are significantly more dangerous than random failures.

Interestingly, increasing the recovery rate ($\gamma = 0.1$ to $\gamma = 0.3$) by a factor of 3 (at $\beta = 0.2$) reduces the peak infection size by nearly 35% (from 6370 to 4113). This suggests that even modest improvements in antivirus response times can disproportionately collapse the epidemic curve, a phenomenon typical of systems operating near the critical threshold λ_c.

Table 5. SIR model results: Complete parameter sweep for infections initiated by super-spreader node 3109.

Rate β	Parameters γ	Final size	Peak size	Time to peak	Avg. infection duration $(1/\gamma)$	Reprod. number (R_0)
0.2	0.1	1.0	6370.2	6.08	10.00	2.00
0.2	0.3	0.0	4113.2	5.20	3.33	0.67
0.2	0.5	0.0	2792.2	4.66	2.00	0.40
0.2	0.7	0.0	1886.2	5.22	1.43	0.29
0.2	0.9	0.0	1349.6	4.61	1.11	0.22
0.4	0.1	0.0	7584.2	3.49	10.00	4.00
0.4	0.3	0.0	5585.6	2.77	3.33	1.33
0.4	0.5	0.0	4550.8	2.42	2.00	0.80
0.4	0.7	0.0	3687.0	2.27	1.43	0.57
0.4	0.9	0.0	2433.6	1.85	1.11	0.44
0.6	0.1	0.6	8203.2	2.79	10.00	6.00
0.6	0.3	0.0	6450.2	2.03	3.33	2.00
0.6	0.5	0.0	5362.4	1.79	2.00	1.20
0.6	0.7	0.0	4647.2	1.66	1.43	0.86
0.6	0.9	0.0	4078.0	1.75	1.11	0.67
0.8	0.1	0.4	8578.4	2.31	10.00	8.00
0.8	0.3	0.0	6942.8	1.59	3.33	2.67
0.8	0.5	0.0	6017.0	1.41	2.00	1.60
0.8	0.7	0.0	5328.2	1.37	1.43	1.14
0.8	0.9	0.0	4727.8	1.22	1.11	0.89
1.0	0.1	0.8	8836.2	2.03	10.00	10.00
1.0	0.3	0.0	7328.2	1.39	3.33	3.33
1.0	0.5	0.0	6437.0	1.17	2.00	2.00
1.0	0.7	0.0	5761.2	1.11	1.43	1.43
1.0	0.9	0.0	5278.0	1.02	1.11	1.11

Table 6. SIR Model Results: Complete parameter sweep for infections initiated by random node 1.

Rate β	Parameters γ	Final size	Peak size	Time to peak	Avg. infection duration $(1/\gamma)$	Reprod. number (R_0)
0.2	0.1	0.4	6455.6	7.38	10.00	2.00
0.2	0.3	0.0	3240.2	4.45	3.33	0.67
0.2	0.5	0.0	2813.4	5.80	2.00	0.40
0.2	0.7	0.0	2008.4	6.26	1.43	0.29
0.2	0.9	0.0	795.6	3.68	1.11	0.22
0.4	0.1	1.0	7621.2	4.27	10.00	4.00
0.4	0.3	0.0	5590.0	3.23	3.33	1.33
0.4	0.5	0.0	4496.4	3.02	2.00	0.80
0.4	0.7	0.0	2215.4	1.90	1.43	0.57
0.4	0.9	0.0	2428.2	2.39	1.11	0.44
0.6	0.1	0.8	8189.2	3.08	10.00	6.00
0.6	0.3	0.0	6411.4	2.43	3.33	2.00
0.6	0.5	0.0	4349.0	1.54	2.00	1.20
0.6	0.7	0.0	4634.6	2.15	1.43	0.86
0.6	0.9	0.0	4119.8	1.99	1.11	0.67
0.8	0.1	0.6	8533.0	2.55	10.00	8.00
0.8	0.3	0.0	6908.4	2.09	3.33	2.67
0.8	0.5	0.0	6031.6	1.81	2.00	1.60
0.8	0.7	0.0	4213.2	1.31	1.43	1.14
0.8	0.9	0.0	3833.8	1.25	1.11	0.89
1.0	0.1	1.4	8852.6	2.25	10.00	10.00
1.0	0.3	0.0	7295.8	1.56	3.33	3.33
1.0	0.5	0.0	6417.2	1.41	2.00	2.00
1.0	0.7	0.0	5728.8	1.31	1.43	1.43
1.0	0.9	0.0	5249.8	1.25	1.11	1.11

To visualize the immediate impact of topology on epidemic initiation, we captured the first 10 steps of the fast_SIR simulation (distinct from time units). **Figures 7** and **8** depict the network state under identical parameters ($\beta = 0.5, \gamma = 0.7$).

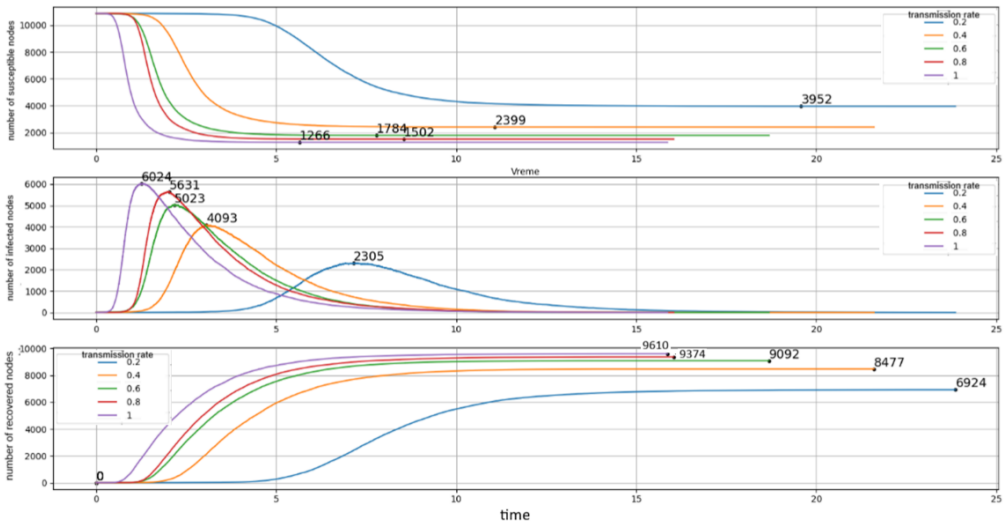

Figure 7. SIR time series with initial infection at super-spreader node 3109.

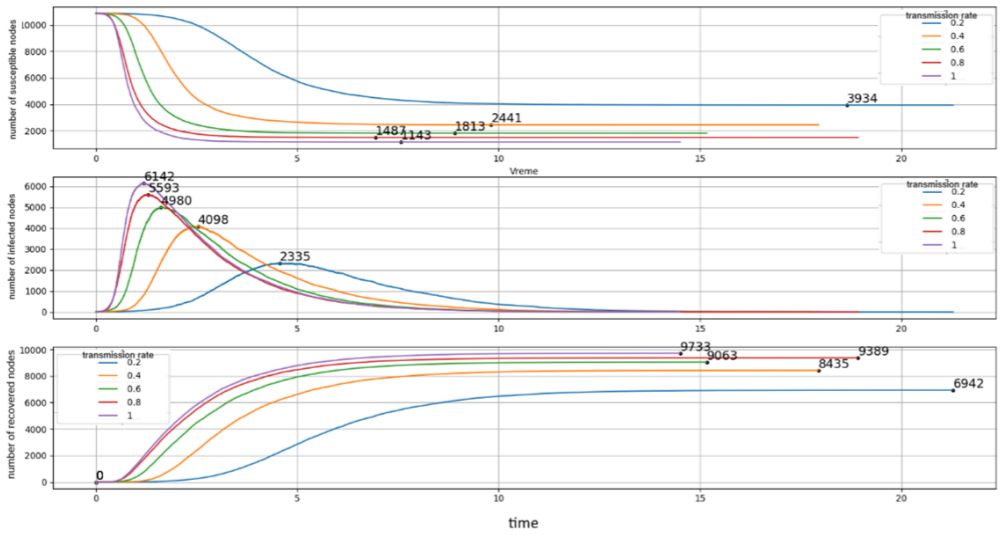

Figure 8. SIR time series with initial infection at random node 1.

In these visualizations, red nodes represent infected peers, blue nodes are susceptible, and green nodes denote identified super-spreaders.

As visualized in **Figure 9**, initiating infection at node 3109 immediately exposes its direct neighbors to the pathogen. In contrast, node 1 (**Figure 10**) exposes only a handful of peers. This topological disparity creates an initial 'shockwave' that drives the system into the exponential growth phase almost instantly, bypassing the slow buildup phase seen in random-node scenarios.

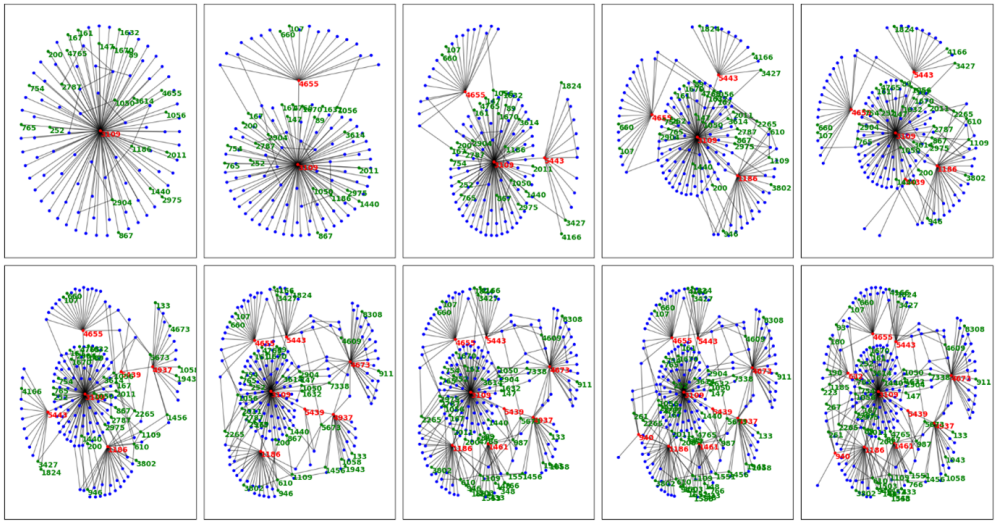

Figure 9. Initial propagation steps (SIR model) starting from a super-spreader (node 3109). The infection rapidly reaches multiple neighborhoods due to the hub's high connectivity.

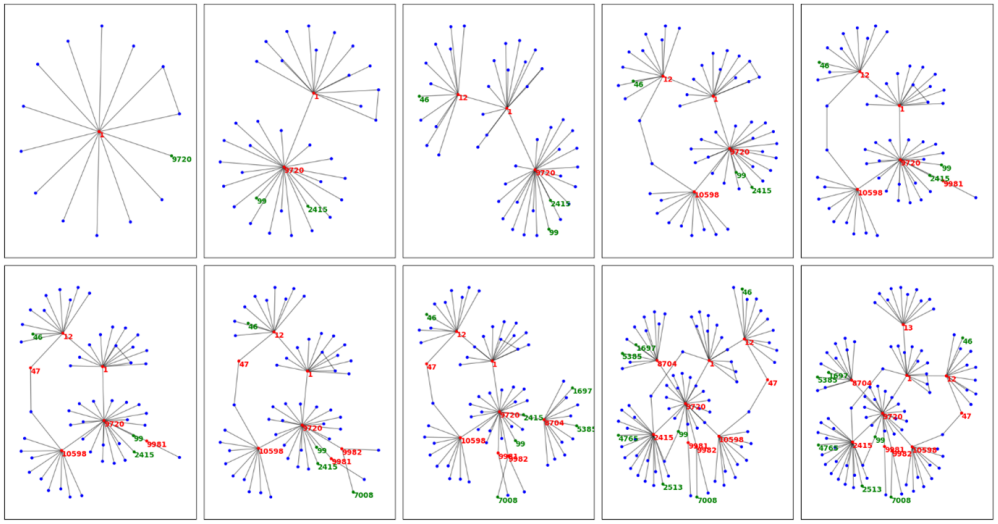

Figure 10. Initial propagation steps (SIR model) starting from a random node (node 1). The infection remains localized within a small cluster.

4. Limitations of classical models and the need for control states

According to the numerical analysis, the classical compartmental model (SI, SIS, SIR) approach gives a fundamental structure for understanding epidemic thresholds; however, in the presence of modern cybersecurity applications, the proposed solution has a serious drawback: The models cannot fully capture the time delays and the active protection mechanisms in digital networks.

4.1. The effect of latency on the dynamics of infections

One major limitation in our simulations of SI and SIS (**Figures 3** and **5**) is the assumption that the infectivity occurs immediately. Biological: This is a disease with a 0 incubation period. Yet the most notable characteristic is that nowadays malware is commonly in a "dormant phase" that acts to bypass detection in order to release its payload [16]. In SEIR (Susceptible-Exposed-Infected-Recovered) models, it was

observed [17] that the lack of an Exposed or Latent (L) compartment causes the propagation velocity to be overestimated. Our simulations show complete saturation of the network ($N = 10,876$) in under 9 time units for $\beta = 1.0$. In a real-world P2P network, however, node handshakes, file transfer protocols, and processing overheads introduce inevitable latency. If we consider a transfer latency of δt, the effective transmission rate is dampened to $\beta' \approx \beta \cdot e^{-\delta t}$. The discrepancy between our numerical "instant" saturation and empirical malware propagation data implies that this latency is non-negligible. Therefore, the results presented here likely represent a theoretical upper bound. Incorporating a latent state L would introduce a necessary time delay τ, effectively flattening the infection curve and providing network administrators a critical window of opportunity for intervention.

4.2. Absence of active defense mechanisms

The standard SIR model assumes that nodes enter the "Recovered" state naturally after infection. In reality, recovery in computer networks is often a result of active countermeasures, such as antivirus software isolating infected files or firewalls blocking malicious traffic. This necessitates a Quarantine (Q) compartment, as proposed in SIQR models [18,19].

Our simulation results for the SIS model (**Table 3**) show that with low recovery rates ($\gamma = 0.1$), the malware becomes endemic. In a real-world scenario, the detection of such a persistent threat would trigger a quarantine response, effectively removing nodes from the infectious pool before they can recover or reinfect others. The lack of a Q-state in our current model explains why the peak infection sizes in our simulations are exceptionally large (up to 96% of the network).

4.3. Transition to SEIR/SIQR

Based on the insight (super-spreader volatility) from Section 3.4, we believe that future attempts to model P2P malware should evolve from simple SIS/SIR architectures to more robust SEIQRS (Susceptible-Exposed-Infected-Quarantined-Recovered-Susceptible). More particularly, the Quarantine system is a key consideration in modeling the effect of "targeted immunization", where super-spreaders (e.g., node 3109) are preferentially isolated. The general equation for this extended model would change the standard SIR model as follows:

$$\frac{dS}{dt} = -\beta SI \tag{8}$$

$$\frac{dE}{dt} = \beta SI - \epsilon E \tag{9}$$

$$\frac{dI}{dt} = \epsilon E - (\gamma + \delta)I \tag{10}$$

$$\frac{dQ}{dt} = \delta I - \eta Q, \tag{11}$$

where ϵ represents the activation rate (Latent $\rightarrow$ Infected), δ is the quarantine rate (Infected $\rightarrow$ Quarantined), and η is the rate of quarantined nodes that get cleaned up and returned to Susceptible or Recovered. Empirical results indicate that without these

variables, standard models fail to predict the effectiveness of cyber-defense tactics. In addition, adding the parameter η renders realistic the reinfection versus permanent patching cycle, a process with particular significance in long-term network security assessment.

4.4. Extending propagation models via fractional derivatives

Expanding standard epidemiological models, such as SIR and SEIR, through the application of fractional-order derivatives represents a highly relevant extension for analyzing propagation dynamics on complex networks [20–22]. Unlike traditional integer-order derivatives, fractional derivatives introduce memory effects and non-local properties, allowing the model to continuously account for the historical states of network nodes during malware propagation. From a mathematically rigorous perspective, the implementation of fractional calculus maintains a robust analytical foundation by utilizing well-defined operators, such as the Caputo derivative [23, 24] or the Mittag-Leffler fractional derivative [25]. These operators alter the differential framework to successfully capture the anomalous diffusion and power-law behaviors inherently found in peer-to-peer networks. Furthermore, adopting fractional epidemiological models significantly improves numerical approximations [26]. The fractional order parameter α acts as an additional degree of freedom, offering enhanced flexibility in parameter estimation and yielding much more accurate numerical solutions and computational stability when modeling complex epidemic curves [23,26].

5. Conclusion

Using the Gnutella dataset, we conducted quantitative analysis on malware propagation dynamics on unstructured peer-to-peer networks. We examined the role of network topology (the impact of high-degree super-spreaders) on the velocity and scale of digital epidemics through systematic application of SI, SIS, and SIR compartmental models. In summary, based on our results, we come to three important conclusions:

1. Due to the scale-free characteristics of P2P networks, they are inherently vulnerable to targeted attacks. Epidemic scenarios that start from a single super-spreader (node 3109) reach saturation 5 times faster than those generated from random nodes were simulated. It confirms that degree centrality is one of the key predictors of epidemic risk in decentralized systems.
2. Classical SI, SIS and SIR models provide excellent predictions of general threshold behavior but suffer from fundamentally overestimating the speed of infection, as they completely ignore the latency time of contemporary malware. The instantaneous transmission expected from SI model ($t_{peak} \approx 9$ units) is a theoretical worst case and not an operational reality.
3. Since the typical SIR model does not naturally dampen the epidemic without high recovery rates, Q states are required as an extra effort to mitigate these effects. We demonstrate that effective cyber defense is not only about recovery (γ), but requires maintaining active isolation of the breached nodes (δ) before they can infect their many neighbors.

Author contributions: Conceptualization, DD and AM; methodology, DD and AM; software, AM; validation, DD and AM; formal analysis, DD and AM; investigation, AM; writing—original draft preparation, AM; writing—review and editing, DD; supervision, DD. Both authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Faculty of Organizational Sciences, University of Belgrade, and in part, by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia through institutional funding (grant number: 200151).

Institutional review board statement: Not applicable.

Informed consent statement: Not applicable.

Data availability statement: The dataset used in this study is publicly available from the Stanford Network Analysis Project (SNAP) repository as the Gnutella peer-to-peer network dataset (August 4, 2002).

Conflict of interest: The authors declare no conflict of interest.

References

1. Cohen F. Computer viruses: theory and experiments. Computers & Security. 1987; 6(1): 22–35.
2. Kephart JO, White SR. Directed-graph epidemiological models of computer viruses. In: Computation: The Micro and the Macro View. World Scientific; 1991. pp. 71–102.
3. Kermack WO, McKendrick AG. 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. 1927; 115(772): 700–721.
4. Newman ME. Spread of epidemic disease on networks. Physical review E. 2002; 66(1): 016128.
5. Martínez Martínez I, Florián Quitián A, Díaz-López D, et al. Malseirs: forecasting malware spread based on compartmental models in epidemiology. Complexity. 2021; 2021(1): 5415724.
6. Lua EK, Crowcroft J, Pias M, et al. A survey and comparison of peer-to-peer overlay network schemes. IEEE Communications Surveys & Tutorials. 2005; 7(2): 72–93.
7. Barabási A-L, Albert R. Emergence of scaling in random networks. Science. 1999; 286(5439): 509–512.
8. Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. Physical review letters. 2001; 86(14): 3200.
9. Ripeanu M, Foster I, Iamnitchi A. Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. arXiv preprint. 2002; arXiv:cs/0209028.
10. Alsubaie AA, Aljoufi MD, Alotaibi AG, et al. Adomian's method for solving a nonlinear epidemic model. Advances in Differential Equations and Control Processes. 2024; 31(1): 95–107.
11. Pavlenko EY, Samarin N. Artificial Immunization in Hierarchical and Peer-to-Peer Networks to Protect Against Cyber Threats. Automatic Control and Computer Sciences. 2025; 59(8): 1519–1526.
12. Kazem H, El Madhoun N, Bouzefrane S, et al. Security challenges and countermeasures in blockchain's peer-to-peer architecture. In: IFIP International Conference on Information Security Theory and Practice. Springer; 2024. pp. 111–127.
13. Adamic LA, Lukose RM, Puniyani AR, et al. Search in power-law networks. Physical Review E. 2001; 64(4): 046135.
14. Cohen R, Havlin S. Scale-free networks are ultrasmall. Physical Review Letters. 2003; 90(5): 058701.
15. Miller JC, Ting T. Eon (epidemics on networks): a fast, flexible Python package for simulation, analytic approximation, and analysis of epidemics on networks. arXiv preprint. 2020; arXiv:2001.02436.
16. Szor P. The Art of Computer Virus Research and Defense. Pearson Education; 2005.
17. Hosseini S, Azgomi MA, Rahmani AT. Malware propagation modeling considering software diversity and

immunization. Journal of Computational Science. 2016; 13: 49–67.
18. Fahreza FR, Hasan M, Santoso KA. Chaotic Outbreak in Discrete Epidemic Model with Vaccination and Quarantine Interventions and Limited Medical Resources. Majalah Ilmiah Matematika dan Statistika. 2025; 25(1): 22–35.
19. Piqueira JRC, Batistela CM. Considering quarantine in the SIRA malware propagation model. Mathematical Problems in Engineering. 2019; 2019(1): 6467104.
20. Nisar KS, Farman M, Abdel-Aty M, et al. A review on epidemic models in sight of fractional calculus. Alexandria Engineering Journal. 2023; 75: 81–113.
21. Nisar KS, Farman M, Abdel-Aty M, et al. A review of fractional order epidemic models for life sciences problems: Past, present and future. Alexandria Engineering Journal. 2024; 95: 283–305.
22. Chen Y, Liu F, Yu Q, et al. Review of fractional epidemic models. Applied Mathematical Modelling. 2021; 97: 281–307.
23. Alshammari NA, Alharthi N, Mohammed Saeed A, et al. Numerical solutions of a fractional order SEIR epidemic model of measles under Caputo fractional derivative. PloS One. 2025; 20(5): e0321089.
24. Gkrekas N. Applying Laplace transformation on epidemiological models as Caputo derivatives. Mathematical Biology and Bioinformatics. 2024; 19(1): 61–76.
25. Sene N. SIR epidemic model with Mittag–Leffler fractional derivative. Chaos, Solitons & Fractals. 2020; 137: 109833.
26. Balzotti C, D'Ovidio M, Lai AC, et al. Effects of fractional derivatives with different orders in SIS epidemic models. Computation. 2021; 9(8): 89.