Chaos, Solitons and Fractals 202 (2026) 117589

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

journal homepage: www.elsevier.com/locate/chaos

Dynamics of defensive and malicious worm co-propagation across networked systems

Andreia Sofia Teixeira [a,b], Ignacio Echegoyen [c], Rasha Shanaz [d], Alberto Aleta [e,*]

[a] BRAN Lab, Network Science Institute, Northeastern University London, London, United Kingdom
[b] LASIGE, Departamento de Informática, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal
[c] Grupo Interdisciplinar de Sistemas Complejos (GISC) & Department of Psychology, Comillas Pontifical University, Madrid, Spain
[d] Department of Physics, Bharathidasan University, Tiruchirappalli, India
[e] Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, Zaragoza, Spain

ARTICLE INFO

Dataset link: https://github.com/aaleta/whiteworms

Keywords:
Complex networks
Black worms
Network epidemiology
White worms
IoT

ABSTRACT

The proliferation of Internet of Things (IoT) devices has greatly enhanced global connectivity but has also amplified cybersecurity risks, particularly from self-propagating malware or black worms. As a countermeasure, some researchers have proposed white worms: benign, self-replicating agents designed to autonomously patch vulnerable systems. Yet, their autonomous behavior raises complex ethical and legal concerns. In this paper, we develop a dynamical model of interacting black and white worms using tools from network epidemiology to explore their co-propagation and emergent behavior across IoT networks. We investigate how parameters related to user response, worm aggressiveness, and network topology shape the system's stability and their dynamics. Our results show that ethical restrictions, such as reduced autonomy or shorter activity, significantly limit the ability of white worms to suppress botnets. Moreover, network structure plays a decisive role in shaping these outcomes. Overall, the study highlights a fundamental tension between ethical design and practical efficacy: to be truly effective, a white worm must behave in ways that challenge its ethical intent.

1. Introduction

The Internet of Things (IoT) has become ubiquitous, with estimates suggesting between two and three connected devices per person worldwide [1]. Even seemingly mundane household items, such as light bulbs and toasters, are now regularly networked. However, this widespread connectivity has led to growing cybersecurity concerns, particularly regarding the escalating threat of malware infections [2]. IoT devices are especially vulnerable due to their limited built-in security and the challenges of deploying timely updates across distributed systems [3]. A striking example of such vulnerability was the 2016 Mirai worm attack, which exploited weakly secured IoT devices to create a large-scale botnet. This botnet was then used to launch a distributed denial-of-service (DDoS) attack against a major DNS provider, temporarily rendering popular websites such as Amazon, Twitter, Netflix, and GitHub inaccessible [4,5].

Such attacks have led to a growing body of research into detecting compromised IoT devices, often leveraging machine learning and neural network approaches [6,7]. Yet, some researchers advocate for more proactive defense mechanisms that move beyond passive detection and response [8–10]. One such concept is the "white worm", a self-propagating agent designed to patch affected systems. A notable example is AntioBioTic, a proposed white worm designed to exploit the same vulnerabilities targeted by malware

[*] Corresponding author.
E-mail address: aaleta@unizar.es (A. Aleta).

https://doi.org/10.1016/j.chaos.2025.117589
Received 30 October 2025; Received in revised form 8 November 2025; Accepted 8 November 2025
Available online 18 November 2025

in order to propagate, patch affected systems, and notify device owners of the breach [11]. Although not the first of its kind [12], AntioBioTic opens an important debate on the ethical and legal implications of deploying self-spreading defensive software [13]. For instance, the European Directive on attacks against information systems explicitly criminalizes the creation of botnets, defined as "establishing remote control over a significant number of computers by infecting them with malicious software" [14]. Even if white worms are arguably not malicious in intent, hijacking third-party devices conflicts with Articles 3 (unauthorized access) and 5 (unauthorized modification) of the Directive.

In this paper, we investigate whether incorporating ethical constraints into white worms affects their defensive effectiveness. Drawing inspiration from AntioBioTic's design, our analysis focuses on their operational viability under ethically bounded behavior, not on the normative or legal dimensions of white worms. Specifically, we explore the conditions under which a constrained white worm could interfere with botnet formation and how such constraints shape co-propagation dynamics. To this end, we model the simultaneous spread of a generic "black worm" and a "white worm" using tools from network epidemiology [15,16]. These methods have a long history in modeling computer virus propagation [17,18], and indeed, several foundational results in network epidemiology originated from studies of digital infections [19].

In this context, a network is an abstraction that encodes the information on the direct connections between devices, not the infrastructure and software that build up the internet. Mathematically, these networks are described by graphs, where nodes represent the entities of interest — IoT devices in our case — and links their direct connections. Note also that even though early malware such as Mirai used a centralized distribution approach, new malware such as IoT_Reaper may spread through peer-to-peer networks [5].

Most existing research on malware propagation in computer networks either focuses on highly detailed, system-specific simulations or on evaluating defensive mechanisms under realistic assumptions [20–28]. Sophisticated simulators can capture the interplay of diverse vulnerabilities, countermeasures, and network topologies [7]. However, our goal is different as we aim to isolate and understand the impact of ethical design constraints on white worm dynamics. For this purpose, we adopt a simplified modeling approach that prioritizes analytical clarity.

Beyond their immediate cybersecurity relevance, the interaction between defensive and malicious worms constitutes a good example of coupled contagion processes in complex systems [29–32]. Such highly interconnected systems are inherently vulnerable to large-scale failures and systemic dysfunction, which may be even triggered by initially localized events [33–36]. Framing malware and white worm co-propagation within this broader context highlights their nature as competing spreading processes on networks, where topology and interaction dynamics jointly determine collective outcomes.

The remainder of this paper is organized as follows. Section 2 presents the model and analyzes its main characteristics under the homogeneous mixing approximation. Section 3 investigates how network topology influences spreading dynamics and the effectiveness of white worm strategies. Finally, Section 4 discusses the implications of our findings and summarizes the main conclusions. Additional analyses and clarifications are provided in Appendix.

2. Model definition and analysis

2.1. Overview of the contagion process

We consider the propagation of two different worms within a network of vulnerable devices (V): a malicious black worm and a benign white worm. The black worm seeks to infiltrate any unprotected device by exploiting an unspecified security vulnerability. Once inside, it propagates through the network to other devices at a transmission rate β_B, awaiting orders from a botmaster to initiate an attack. The set of devices infected by the black worm is denoted by B.

Conversely, the white worm aims to spread across the network to secure vulnerable devices. It does so with a transmission rate β_W. Following the approach proposed in [11], we assume that the white worm attempts to behave ethically. Although it breaches the system, initially it only notifies the device owner of the issue and then enters a dormant state (D), consuming no resources and making no modifications to the system.

Upon notification, owners may patch their devices at rate γ. When this occurs, the white worm detects the update and, having fulfilled its defensive purpose, erases itself. However, many IoT devices cannot easily notify their owners-or owners may lack the motivation or technical ability to patch them. Consequently, the white worm may reactivate from its dormant state at rate ϵ, entering an active state (W). While active, it propagates to other devices for a limited period before patching the system and deleting itself. This final transition is governed by rate μ.

Note that the presence of a white worm does not automatically eliminate the device's vulnerability as doing so would require system modification, which the ethical design avoids except as a last resort. As a result, both worms may temporarily coexist on the same device. The possible states and transitions of the model are illustrated in Fig. 1 and summarized in Table 1.

Our analysis focuses on the two parameters that control the ethical behavior of the white worm: γ, representing the rate at which owners secure their devices after being notified; and ϵ, the inverse of the time the white worm allows owners to act before it intervenes directly. However, we recognize that ethical considerations cannot be reduced to only these two parameters. Thus, in the rest of the paper, we will often discuss these parameters using the term "roughness" as it will be explained in the next section. In addition, we explore how the relative propagation speed — the ratio between β_W and β_B — influences the overall dynamics. In particular, we examine whether a white worm spreading at the same rate, more slowly, or faster than its malicious counterpart can effectively prevent large-scale infection.

Table 1
List of symbols and their explanations.

Symbol	Explanation
V	Number of vulnerable devices
B	Devices infected by the black worm
D	Devices infected by the white worm in a dormant state
W	Devices infected by the white worm in an active state
DB	Devices infected by both worms, with the white worm dormant
WB	Devices infected by both worms, with the white worm active
P	Devices fully protected after patching
β_B	Infection rate of the black worm
β_W	Infection rate of the white worm
ϵ	Inverse of the time allowed for the owner to patch the device
μ	Inverse of the active spreading duration of the white worm
γ	Rate at which owners act upon receiving a notification
ρ^X	Fraction of devices in state X
ϕ^X	Total fraction of devices carrying worm X (black or white) in an active state

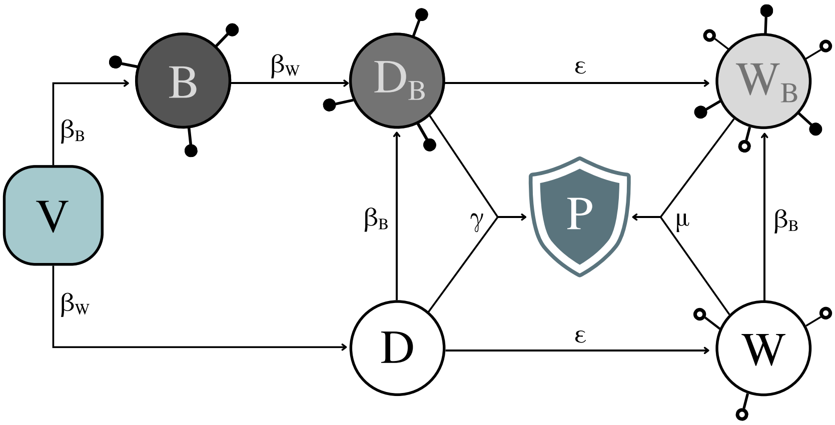

Fig. 1. Model scheme - Model describing the simultaneous spread of black and white worms. Vulnerable devices (V) can be infected by either worm. Devices infected by the black worm (B) actively spread it. Devices infected by the white worm enter a dormant state (D) until a user patch or self-activation occurs. Activated white worms (W) propagate briefly before patching and securing the system. Devices in any state with a white worm may also be infected by a black worm (DB, WB), and vice versa. Once a device is updated, either by user action or white worm intervention, it becomes protected (P), closing its vulnerability.

2.2. Dynamics under the homogeneous mixing approximation

As it is commonly done in the study of malware propagation [20,21,28], to gain insight on the dynamical properties of this model, we begin with the homogeneous mixing approximation. This approximation considers that all devices can potentially interact every other device. Consequently, malware propagation depends only on the overall fractions of vulnerable and infected devices, not on network topology.

Let $\rho^X(t)$ denote the fraction of devices in state X at time t. The dynamics of the system are then governed by the following set of differential equations:

$$\begin{aligned}
\dot{\rho}^V &= -\left(\beta_B\phi^B + \beta_W\phi^W\right)\rho^V \\
\dot{\rho}^B &= \beta_B\phi^B\rho^V - \beta_W\phi^W\rho^B \\
\dot{\rho}^D &= \beta_W\phi^W\rho^V - \left(\beta_B\phi^B + \epsilon + \gamma\right)\rho^D \\
\dot{\rho}^{DB} &= \beta_B\phi^B\rho^D + \beta_W\phi^W\rho^B - (\epsilon+\gamma)\rho^{DB} \\
\dot{\rho}^W &= \epsilon\rho^D - \left(\beta_B\phi^B + \mu\right)\rho^W \\
\dot{\rho}^{WB} &= \epsilon\rho^{DB} + \beta_B\phi^B\rho^W - \mu\rho^{WB} \\
\dot{\rho}^P &= \mu\left(\rho^{WB} + \rho^W\right) + \gamma\left(\rho^{DB} + \rho^D\right)
\end{aligned} \tag{1}$$

where $\phi^B = \rho^B + \rho^{DB} + \rho^{WB}$ is the total fraction of devices carrying an active black worm, and $\phi^W = \rho^W + \rho^{WB}$ represents the corresponding quantity for the white worm. By construction, $\sum\rho^X = 1$. For the definition of all parameters and states, see Section 2.1 and Fig. 1. Following standard practice, we set $\mu = 1$ without loss of generality (as time can be rescaled accordingly) though we retain it in analytical expressions for completeness.

A key quantity in epidemic modeling is the basic reproduction number (R_0), which measures the expected number of secondary infections generated by a single infected entity in a fully susceptible population [37]. For the black worm, which follows a simple SI-like propagation pattern, the basic reproduction number is

$$R_0^B = \beta_B \tag{2}$$

For the white worm, we derive the equivalent quantity using the next-generation matrix approach (see Appendix A for details):

$$R_0^W = \frac{\beta_W}{\mu} \cdot \frac{\epsilon/\gamma}{\epsilon/\gamma + 1} \equiv \frac{\beta_W}{\mu} \cdot \frac{\epsilon'}{\epsilon' + 1}. \tag{3}$$

Hence, the infectiveness of the white worm depends on the ratio ϵ/γ, which we define as the roughness of the worm, ϵ'. This ratio encapsulates the ethical trade-off in the worm's design:

- When $\epsilon' = 0$ (either $\epsilon = 0$ or $\gamma \to \infty$), the worm never self-activates or owners respond instantaneously, resulting in an almost passive agent that only notifies.
- When $\epsilon' \to \infty$ (either $\epsilon \to \infty$ or $\gamma = 0$), the worm always activates without allowing owners to intervene, representing the most aggressive configuration.

Thus, ϵ' quantifies how forcefully the white worm enforces protection: higher values correspond to more coercive, "rougher" behavior. As we will show, many macroscopic outcomes depend solely on this ratio.

Using R_0^W, the final fraction of protected devices can be estimated from the transcendental equation [38]:

$$\rho^P(t \to \infty) = 1 - e^{-R_0^W \rho^P}. \tag{4}$$

Fig. 2a compares the analytical solution of Eq. (4) with numerical integrations of the full system in Eq. (1). As expected, the fraction of protected devices increases with the white worm's infection rate β_W. Importantly, protection also grows with the worm's roughness ϵ'. Rougher worms protect more devices even at moderate transmission rates, whereas ethically restrained worms require significantly higher β_W to achieve comparable protection levels.

To determine whether protection arises primarily from owner intervention or forced updates, Fig. 2b compares the proportion of devices patched by their owners and the total amount of protected devices. When $\epsilon' < 1$ ($\epsilon < \gamma$), most devices are protected voluntarily by their owners, but overall coverage remains limited. Larger-scale protection requires higher roughness, at the cost of increased autonomous intervention, or higher transmission rates.

We now examine the interplay between the black and white worms as they spread simultaneously. Linearizing the system around $t \to 0$ [39], we find that a botnet can grow only if:

$$\beta_B > \frac{1}{2}\left[-(\epsilon + \gamma + \mu) + \sqrt{(\epsilon + \gamma - \mu)^2 + 4\beta_W \epsilon}\right]. \tag{5}$$

Note that this expression depends on ϵ and γ separately and not on their ratio (see Appendix B for the complete derivation).

Fig. 2c shows the final botnet size as a function of β_W and β_B, with the theoretical threshold given by the dashed line from Eq. (5). The agreement between theory and simulation is excellent: below the threshold, the botnet size approaches zero. Moreover, in the limit $\epsilon \to \infty$, Eq. (5) simplifies to $\beta_B > \beta_W - \mu$, implying that even a maximally rough white worm must spread faster than $\beta_B + \mu$ to completely prevent botnet formation (see also Fig. B.1 in Appendix B).

Finally, we study the duration for which the botnet remains active. Fig. 2d shows the number of time steps during which the botnet exceeds a given size threshold. For instance, the botnet occupies more than 25% of devices for nearly the entire simulation, limiting protection to about 75%. The dashed line, corresponding to Eq. (4), accurately predicts this upper bound. For larger botnets, we observe that the white worm eventually dismantles them provided its roughness exceeds a critical threshold. Appendix C further explores the dependence of this threshold on β_W and γ.

3. Spreading on networks

3.1. Network topologies

The topology of the network through which a contagion spreads has a major impact on its dynamics [18,19,40]. Therefore, understanding how different network structures affect the behavior of our model is essential. In real-world computer and IoT systems, network structures are rarely homogeneous. Further, in the literature on malware spreading and IoT systems, it is possible to find many different topologies that depend both on the communication scheme and the actual infrastructure building up the network.

Common examples include simple configurations such as rings, stars, or meshes for small networks, and larger-scale structures such as small-world, scale-free (SF), Erdős–Rényi (ER), or random geometric graphs [7]. For instance, networks of mobile wearable devices communicating through wireless technologies are better represented by ER or random geometric graphs [41], while non–peer-to-peer systems with centralized servers would follow more hierarchical communication patterns.

For simplicity, here we focus on ER and SF networks—two canonical topologies widely studied in network epidemiology that represent limiting cases of homogeneous and highly heterogeneous connectivity, respectively. These choices allow us to examine the influence of network structure in a controlled way. Equivalent qualitative results were also obtained for hierarchical networks generated using a stochastic block model (not shown). All networks consist of 10^4 nodes with an average degree of $\langle k \rangle = 10$, generated using the Python package NetworkX (version 3.1) [42].

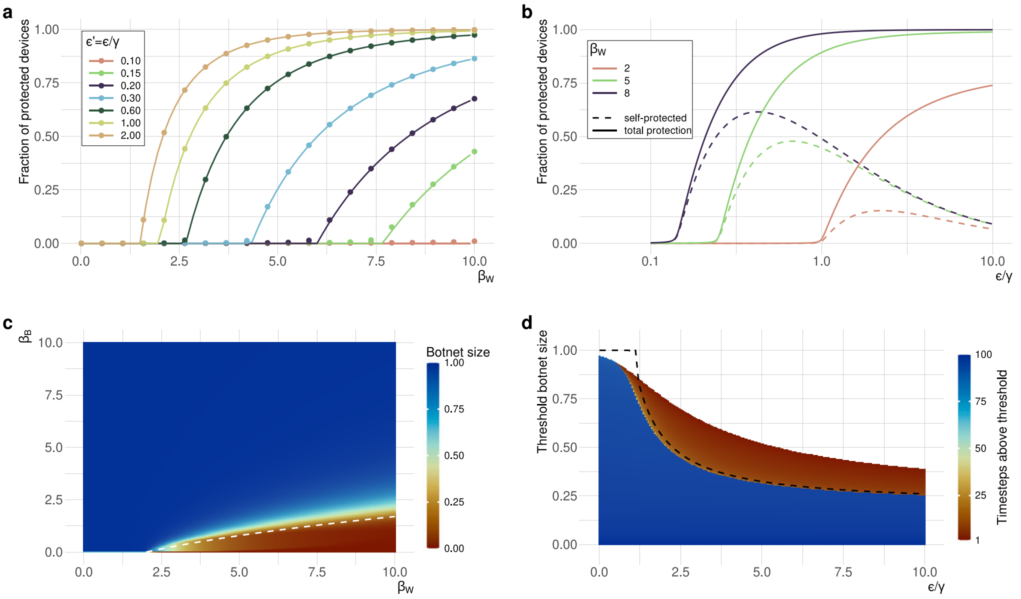

Fig. 2. Dynamics of the model under the homogeneous mixing approximation - (a) Final fraction of protected devices as a function of the white worm infection rate, β_W, and its roughness, $\epsilon' = \epsilon/\gamma$. A higher roughness yields greater protection even for low β_W. Dots correspond to numerical integration of Eq. (1), while lines show the analytical solution from Eq. (4). **(b)** Fraction of protected devices as a function of roughness ϵ' for selected values of β_W. Solid lines represent the total number of protected devices, while dashed lines indicate those protected by their owner. The fraction of devices forcefully protected is the difference between the two. When $\epsilon < \gamma$, most devices are self-protected but overall protection is low; for higher roughness, coverage increases mainly through forced protection. **(c)** Final botnet size as a function of β_W and β_B, with $\epsilon' = \epsilon = \gamma = 1$. The dashed line denotes the theoretical threshold given by Eq. (5). Under these conditions, the white worm must spread faster than the black worm to prevent botnet formation. **(d)** Duration (in timesteps) for which the botnet size exceeds a given threshold, for $\beta_B = 1$ and $\beta_W = 2$. The dashed line corresponds to the theoretical prediction from Eq. (4). For a fixed botnet size, increasing roughness shortens the botnet's lifetime.

3.2. Stochastic simulations

Stochastic simulations are particularly important for discrete systems where finite-size effects are non-negligible. We simulate the joint propagation of both worms using the Gillespie algorithm [43], originally developed for chemical reaction systems and now widely used to model epidemic dynamics in continuous time [38,44]. In particular, we employ the implementation available in the Python package Epidemics on Networks (EoN) (version 1.1) [45].

Each simulation starts with one device infected by the black worm and another by the white worm. Since the process has an absorbing state, the simulation runs until no further transitions are possible—i.e., when the white worm has disappeared and the black worm can no longer spread, either because there are no more vulnerable devices left to infect or because the black worm has been completely eradicated. To ensure statistical robustness, each configuration is simulated 1000 times, and results are averaged. For fair comparison across different network topologies, infection rates are normalized by the average degree, $\langle k \rangle$.

As a validation step, Appendix D shows results obtained using the Gillespie algorithm on a complete network, which converge to those from the homogeneous-mixing approximation in Section 2 for sufficiently large systems.

3.3. Results

We first compute the maximum botnet size in the β_W–β_B plane for both network types to identify parameter regions of interest. As shown in Fig. 3a, the behavior of the model on a random (ER) network closely resembles that obtained under the homogeneous mixing approximation. This similarity is expected, as ER networks lack strong structural correlations. The infection threshold under our model's normalization is $\beta_W > 1$, identical to that of the homogeneous case, since infection rates are normalized by the average degree. Consequently, the botnet size distribution mirrors that in Fig. 2c, and the theoretical prediction from the homogeneous model still provides a good approximation.

In contrast, the SF network has a slightly different behavior. It is well known that in SF networks, the infectivity threshold vanishes as the size of the system grows [19]. In our case, with a network of only 10,000 nodes and an average degree of 10, we are far from that limit. Nonetheless, we observe that the white worm can spread over a larger area of the plane and prevent the black worm from taking control of the whole system, even though we still need large values of β_W to completely protect the system.

Given the similar dependence on β, Fig. 4 explores the same parametrization used for the homogeneous system to enable direct comparison. We first examine the final fraction of protected devices in the ER (Fig. 4a) and SF (Fig. 4b) networks. For the ER

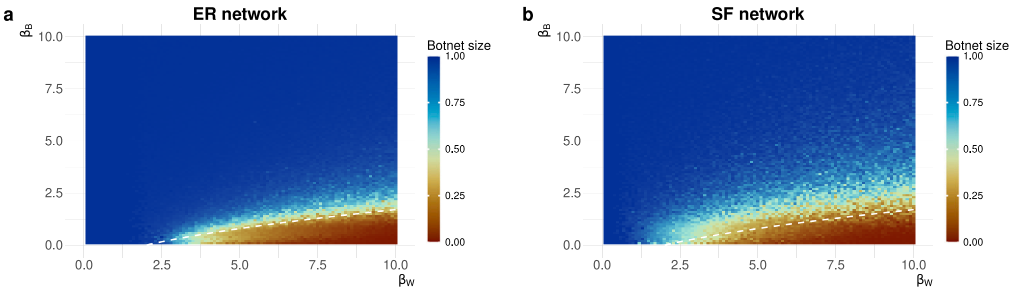

Fig. 3. Final botnet size as a function of the spreading rates - Botnet size at the end of the simulation in the β_W–β_B plane for $\epsilon' = \epsilon = \gamma = 1$. The dashed line indicates the theoretical threshold from the homogeneous mixing approximation. **(a)** Results for the ER network, which closely reproduce the homogeneous case (cf. Fig. 2c). **(b)** Results for the SF network, which exhibits a broader region where the botnet fails to occupy the entire system compared to the homogeneous case.

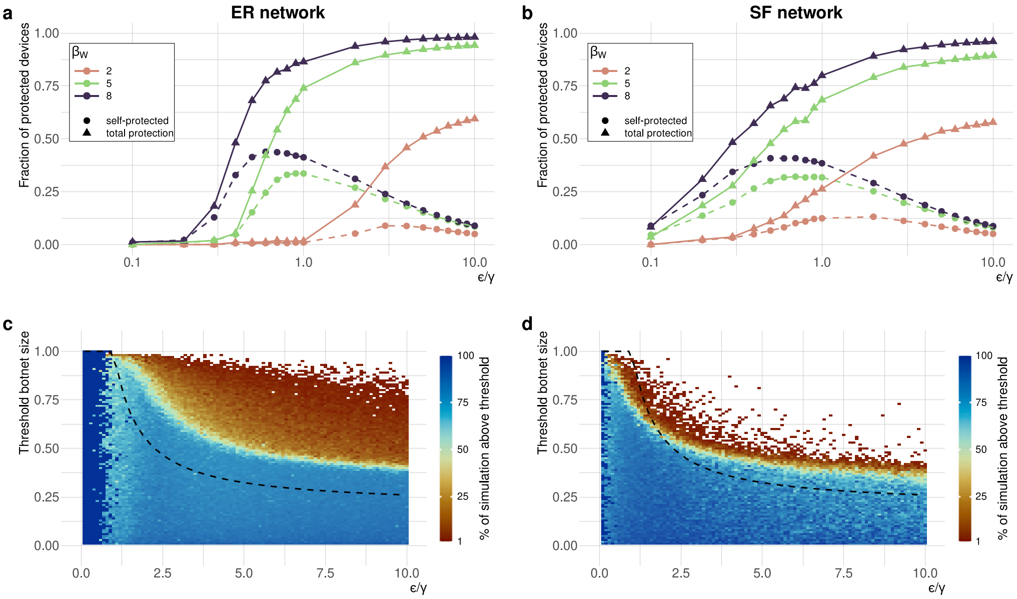

Fig. 4. Protection under different network topologies - The first row shows the fraction of protected devices as a function of the worm's roughness, $\epsilon' = \epsilon/\gamma$, for selected values of β_W and $\beta_B = 1$. Triangles denote the total fraction of protected devices (including both owner- and worm-driven updates), while circles indicate those protected voluntarily by their owners. **(a)** ER network: behavior closely matches the homogeneous case. **(b)** SF network: the white worm spreads more easily even at low ϵ', though total protection remains limited. The second row shows the fraction of the simulation time during which the botnet size exceeds a given threshold. The dashed line marks the theoretical prediction from the homogeneous approximation. **(c)** ER network: the region where the botnet persists for the entire simulation is narrow ($\epsilon' < 1$). **(d)** SF network: only when $\epsilon' \rightarrow 0$ can the botnet occupy the whole system for the full simulation.

case, the results closely follow those obtained under homogeneous mixing, particularly for $\beta_W = 2$. When β_W increases further, deviations arise due to the finite connectivity of nodes: once all neighbors of an infected device become protected, further increases in infectivity no longer enhance spreading efficiency.

The behavior differs slightly in SF networks. The white worm can still reach a macroscopic fraction of the system even at low roughness (ϵ'), but the overall protection coverage tends to be smaller. This reduction arises from the system's discrete nature and from the skewed degree distribution of SF networks: many nodes have degrees below the average, leading to a faster depletion of vulnerable neighbors. Regarding the distinction between forced and owner-initiated protection, the results remain similar to the homogeneous case, as that mechanism is independent of network topology.

Finally, we study the persistence of the botnet over time. Unlike the continuous equations in Eq. (1), which are valid for arbitrarily long times, Gillespie simulations terminate once no further transitions can occur. Therefore, we measure the fraction of the total simulation time during which the botnet exceeds a given size instead of the number of timesteps.

In Fig. 4c, we find that ER topology induces two competing effects. On one hand, the botnet can remain above the theoretical limit predicted by the homogeneous approximation for a significant fraction of the simulation. On the other, the region where the botnet persists throughout the entire simulation is limited to a narrow band of low roughness ($\epsilon' < 1$). A comparable pattern appears

Table 2

Summary of optimal model parameters - For each parameter, we report the values required for the white worm to effectively suppress the black worm. Successful eradication requires the white worm to infect most of the system, which can be achieved either by high infectivity (β_W) or by remaining active for long periods (μ). If users patch their devices too quickly (γ), the worm cannot spread efficiently. Similarly, it should activate rapidly (ϵ) to minimize the chance of being neutralized before propagation.

Description	Parameter	Effective value
Infection rate of white worm	β_W	High
Activation rate of white worm	ϵ	High
Rate of voluntary patching	γ	Low
Rate of self-destruction	μ	Low

for SF networks, although the region exceeding the homogeneous prediction is smaller, and the botnet survives for the full duration only when $\epsilon' \approx 0$.

4. Conclusions

The intersection of cybersecurity and ethics presents complex and often conflicting dilemmas. In this work, we explored one such challenge through the lens of white worms—self-propagating agents proposed as a defensive measure for Internet of Things (IoT) systems. The use of white worms inherently lies on an ethical and legal boundary: although they may strengthen collective cybersecurity, their autonomous propagation and system modification occur without explicit user consent, potentially breaching privacy and existing legislation. Some authors have suggested that their deployment could be regulated or managed by governments [46], yet this introduces its own set of governance and accountability concerns that extend beyond the scope of this study.

Here, rather than engaging directly in the ethical debate, we modeled the dynamics of a white worm inspired by the behavioral principles proposed in the literature [11,13]. We characterized the worm by its roughness, defined as the ratio between its activation rate — representing the intensity of its use of system resources for propagation and forced protection (ϵ) — and the rate at which users voluntarily act upon receiving a security notification (γ). Another relevant parameter is μ, describing how long the white worm remains active before patching the system and deleting itself. Since other parameters can be rescaled as a function of μ, its role can be studied indirectly through variations in β_W, which encapsulates both the worm's infectivity and its active lifetime. Thus, large values of β_W correspond either to faster identification of vulnerable devices or to prolonged activity before self-destruction (see Table 2).

We then examined how different network topologies influence these dynamics. The main conclusion is that although structured networks facilitate partial protection, achieving full system coverage becomes more difficult. Specifically, the region in which the botnet remains active for extended periods is considerably smaller in networks with heterogeneous (power-law) degree distributions. These findings highlight the critical role of network structure in determining the feasibility and efficiency of white worm deployment strategies.

Despite these insights, our study has several limitations. First, it relies on simplifying assumptions about the behavior of both worms and users, which may not fully capture real-world complexities. For example, we assumed that once a device is patched it becomes completely immune to reinfection, whereas in practice updates may not close all vulnerabilities. In some cases, rebooting might remove both worms, leaving the system temporarily unprotected. Moreover, our model does not consider direct interactions between the worms, such as the white worm actively detecting and removing the black worm from infected devices.

Future work could extend the model to incorporate such interactions and recurrent infections, for instance by transforming the SEIR-like formulation into a SEIRS-like model. Empirical validation through experiments or real-world data would also help assess the predictive power of the model and refine its assumptions. Furthermore, studying the dynamics on realistic IoT network topologies or under different communication paradigms such as centralized update servers would provide a more accurate assessment of deployment feasibility and control effectiveness.

In conclusion, our results show that the co-dynamics of black and white worms can generate complex nonlinear behaviors that determine whether a botnet survives or collapses. However, the success of white worms depends on aggressive parameter regimes (high infectivity, prolonged activity, and low voluntary user response) that challenge the very ethical principles motivating their design. Thus, while technically promising, the idea of an ethical white worm remains paradoxical: to be effective, it must behave in ways that are inherently intrusive and ethically questionable. Comparable levels of protection might instead be achieved through safer alternatives, such as coordinated automatic patching or federated defense systems, which emulate the cooperative dynamics of white worms without requiring autonomous propagation.

CRediT authorship contribution statement

Andreia Sofia Teixeira: Writing – review & editing, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Ignacio Echegoyen:** Writing – review & editing, Visualization, Methodology, Investigation. **Rasha Shanaz:** Writing – review & editing, Visualization, Methodology, Investigation. **Alberto Aleta:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was quick-started in the Complexity72 h workshop, held at IFISC in Palma, Spain, 26–30 June 2023, https://www.complexity72h.com. All authors would like to acknowledge the participation of Francesco Bonacina, Diego Escribano, Marcus Krellner and Francesco Paolo Nerini who participated in the preliminary analysis performed during the workshop. AA also thanks Miguel Tarancón for evaluating the viability of the approach in his Bachelor's thesis. RS acknowledges the support of SERB International Travel Support grant ref. ITS/2023/001976. IE acknowledges the financial support of the Unversidad Pontificia Comillas. AST acknowledges support by FCT – Fundação para a Ciência e Tecnologia – through the LASIGE Research Unit, ref. UID/408/2025. AA acknowledges support from the grant RYC2021-033226-I funded by MCIN/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR, and through the grant PID2023-149409NB-I00 funded by MICUI/AEI/10.13039/501100011033 and FEDER, UE. We acknowledge the use of the computational resources of COSNET Lab at Institute BIFI, funded by Banco Santander (grant Santander-UZ 2020/0274) and by the Government of Aragón (grant UZ-164255). The funders had no role in study design, data collection and analysis, the decision to publish, or the preparation of the manuscript.

Appendix A. Derivation of the basic reproduction number, R_0

The next-generation matrix defines R_0 as [37,47]:

$$R_0 = \rho(-FV^{-1}), \tag{A.1}$$

where ρ denotes the spectral radius, and F and V represent, respectively, the rate of appearance of new infections and the rate of transfer of individuals across states. In the case of the white-worm, which follows a SEIR-like structure, we have:

$$F = \begin{pmatrix} 0 & \beta_W \\ 0 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} -(\epsilon+\gamma) & 0 \\ \epsilon & -\mu \end{pmatrix}. \tag{A.2}$$

Hence,

$$FV^{-1} = \begin{pmatrix} \dfrac{\beta_W \epsilon}{(\epsilon+\gamma)\mu} & \dfrac{\beta_W}{\mu} \\ 0 & 0 \end{pmatrix}. \tag{A.3}$$

So that the spectral radius is

$$R_0 = \rho(-FV^{-1}) = \frac{\beta_W}{\mu} \cdot \frac{\epsilon}{\epsilon+\gamma} \tag{A.4}$$

Appendix B. Early growth of the worms

Following [39], we linearize the system of equations around $t \to 0$ to approximate the early dynamics. For the black worm:

$$\dot{\rho}^B \approx \beta_B \rho^B, \tag{B.1}$$

so that in the early phase of an outbreak the size of the botnet will grow as:

$$\rho^B(t) = \rho^B(0)e^{\beta_B t}. \tag{B.2}$$

That is, exponentially with rate β_B.

For the white worm, the early dynamics are governed by:

$$\dot{\rho}^D \approx \beta_W \rho^W - (\epsilon+\gamma)\rho^D \tag{B.3}$$

$$\dot{\rho}^W \approx \epsilon\rho^D - \mu\rho^W. \tag{B.4}$$

In this case, we can obtain the early growth of the white worm outbreak using the eigenvalues of the Jacobian matrix:

$$J = \begin{pmatrix} -(\epsilon+\gamma) & \beta_W \\ \epsilon & -\mu \end{pmatrix}, \tag{B.5}$$

with eigenvalues:

$$\lambda = \frac{tr(J) \pm \sqrt{(tr(J))^2 - 4\det(J)}}{2}. \tag{B.6}$$

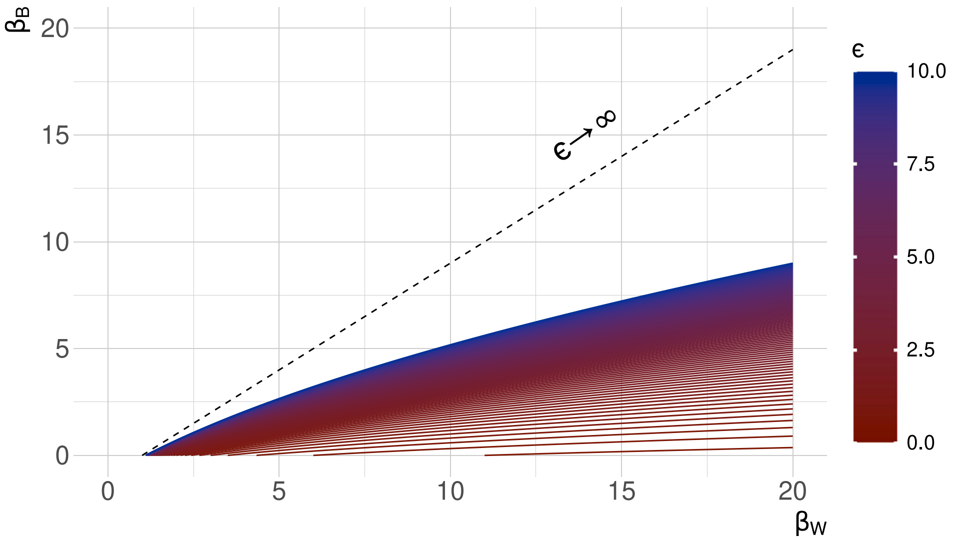

Fig. B.1. Growth of the botnet as a function of ϵ - The plot depicts the critical boundary separating the botnet-free region of the $\beta_W - \beta_B$ plane. Below the line, the botnet cannot grow.

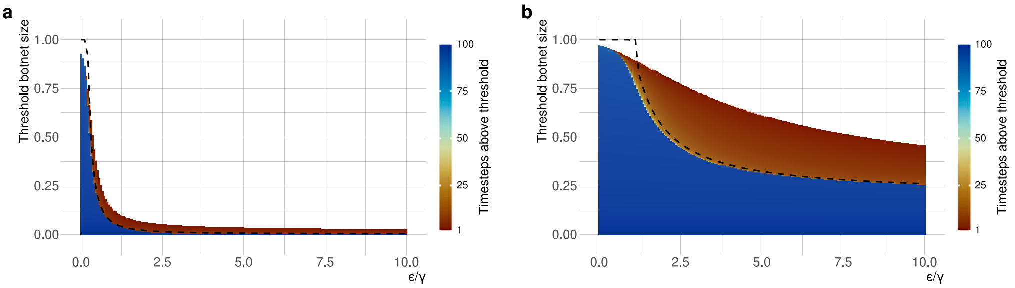

Fig. C.1. Threshold botnet size - **(a)** number of timesteps during which the botnet remains above a given size for a more infectious white worm ($\beta_B = 1$, $\beta_W = 6$). **(b)** same setup as in the main text ($\beta_B = 1$, $\beta_W = 2$) but with $\gamma = 0.5$.

Hence, the leading eigenvalue, which determines the early growth rate of the white worm, is:

$$\lambda = \frac{-(\epsilon+\gamma+\mu)+\sqrt{(\epsilon+\gamma-\mu)^2+4\beta_w\epsilon}}{2}. \tag{B.7}$$

Comparing both expressions, we can set an approximate threshold for the black worm to be able to spread through the system, since if the white worm spreads much faster it may remove it completely before it can grow. As such, the black worm will only be able to spread substantially if

$$\beta_B > \frac{1}{2}\left[-(\epsilon+\gamma+\mu)+\sqrt{(\epsilon+\gamma-\mu)^2+4\beta_W\epsilon}\right]. \tag{B.8}$$

Fig. B.1 shows this threshold as a function of ϵ for $\gamma = 1$. In the limit $\epsilon \to \infty$, the right-hand side approaches $\beta_W - \mu$, implying that for the white worm to completely suppress the botnet, its effective spreading rate must satisfy $\beta_W > \beta_B + \mu$.

Appendix C. Threshold botnet size

The black worm will eventually cover the whole system, since it never deletes itself unless the white worm protects the device. Thus, the maximum size of the botnet depends directly on the total coverage achieved by the white worm, as given by Eq. (4). The only effective way of reducing the botnet size is therefore to increase R_0^W.

As shown in Fig. C.1a, where we set $\beta_B = 1$ and $\beta_W = 6$, this is indeed the case: the botnet remains large only when the white worm roughness is close to zero. Fig. C.1b illustrates a similar situation with the same parameters as in the main text, but with $\gamma = 0.5$. In this case, the theoretical threshold is still valid, although it takes slightly longer for the white worm to eradicate the botnet.

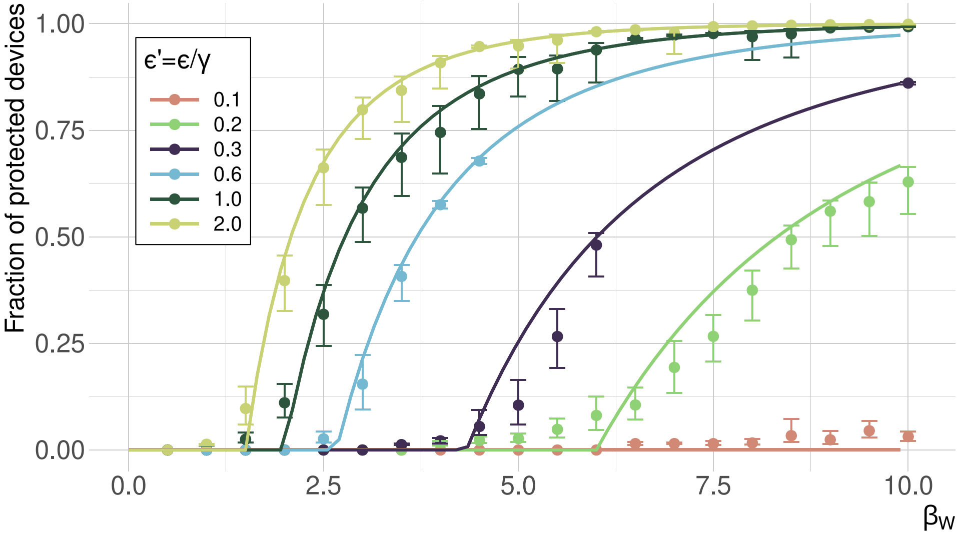

Fig. D.1. Simulations on the complete graph - Fraction of protected devices at the end of the simulation as a function of β_W for several values of the roughness $\epsilon' = \epsilon/\gamma$. Dots represent the average of 1000 simulations, and error bars denote the 95% confidence interval. Lines show the theoretical prediction from the homogeneous system (Eq. (4)).

Appendix D. Stochastic simulation on the complete graph

The homogeneous mixing approximation assumes that each device can communicate directly with every other device. We can mimic this condition in a network by connecting all nodes to each other, forming a complete graph. To verify that the stochastic model accurately reproduces the dynamics described by Eq. (1), Fig. D.1 compares the results of stochastic simulations on a complete graph of 10,000 nodes with the theoretical predictions from Eq. (4).

The agreement between both approaches is very good, with only minor deviations due to finite-size effects and stochastic fluctuations. Some additional noise is expected because of the limited number of simulation runs (1000), constrained by computational cost.

Data availability

The code used in this study is publicly available at https://github.com/aaleta/whiteworms.

References

[1] Cisco annual internet report (2018–2023). 2022, White Paper, URL https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html.

[2] Sengupta J, Ruj S, Das Bit S. A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoT. J Netw Comput Appl 2020;149:102481. http://dx.doi.org/10.1016/j.jnca.2019.102481.

[3] Kaur B, Dadkhah S, Shoeleh F, Neto ECP, Xiong P, Iqbal S, et al. Internet of Things (IoT) security dataset evolution: Challenges and future directions. Internet Things 2023;22:100780. http://dx.doi.org/10.1016/j.iot.2023.100780.

[4] Sinanović H, Mrdovic S. Analysis of Mirai malicious software. In: 2017 25th international conference on software, telecommunications and computer networks. IEEE; 2017, p. 21–3. http://dx.doi.org/10.23919/SOFTCOM.2017.8115504.

[5] Vlajic N, Zhou D. IoT as a land of opportunity for DDoS hackers. Computer 2018;51(7):26–34. http://dx.doi.org/10.1109/MC.2018.3011046.

[6] Wei C, Xie G, Diao Z. A lightweight deep learning framework for botnet detecting at the IoT edge. Comput Secur 2023;129:103195. http://dx.doi.org/10.1016/j.cose.2023.103195.

[7] Chee KO, Ge M, Bai G, Kim DD. IoTSecSim: A framework for modelling and simulation of security in Internet of things. Comput Secur 2024;136:103534. http://dx.doi.org/10.1016/j.cose.2023.103534.

[8] Lu W, Xu S, Yi X. Optimizing active cyber defense. In: Decision and game theory for security. Cham, Switzerland: Springer; 2013, p. 206–25. http://dx.doi.org/10.1007/978-3-319-02786-9_13.

[9] Yamaguchi S. Botnet defense system: Concept, design, and basic strategy. Information 2020;11(11):516. http://dx.doi.org/10.3390/info11110516.

[10] Yamamoto Y, Fukushima A, Yamaguchi S. Implementation of white-hat worms using mirai source code and its optimization through parameter tuning. Futur Internet 2024;16(9):336. http://dx.doi.org/10.3390/fi16090336.

[11] De Donno M, Dragoni N, Giaretta A, Mazzara M. Antibiotic: Protecting IoT Devices Against DDoS Attacks. In: Proceedings of 5th international conference in software engineering for defence applications. Cham, Switzerland: Springer; 2018, p. 59–72. http://dx.doi.org/10.1007/978-3-319-70578-1_7.

[12] Ferronato G. IoT white worms : design and application [Ph.D. thesis], University of Twente; 2020, URL https://essay.utwente.nl/83003.

[13] De Donno M, Felipe JMD, Dragoni N. ANTIBIOTIC 2.0: A fog-based anti-malware for Internet of Things. In: 2019 IEEE European symposium on security and privacy workshops. IEEE; 2019, p. 17–9. http://dx.doi.org/10.1109/EuroSPW.2019.00008.

[14] Directive 2013/40/EU of the European Parliament and of the Council of 12 2013 on attacks against information systems and replacing Council Framework Decision 2005/222/JHA. 2013, URL https://eur-lex.europa.eu/eli/dir/2013/40/oj.

[15] Tong L, Zhang S, Sun H, Cai Y, Zhang J, Li J. Advancements in epidemic transmission suppression: A comprehensive survey. IEEE Trans Netw Sci Eng 2025;1–21. http://dx.doi.org/10.1109/TNSE.2025.3579136.

[16] Liu Y, Liang G, Wang X, Zhu P, Wang Z. Diffusion containment in complex networks through collective influence of connections. IEEE Trans Inf Forensics Secur 2023;19:1510–24. http://dx.doi.org/10.1109/TIFS.2023.3338423.

[17] Murray WH. The application of epidemiology to computer viruses. Comput Secur 1988;7(2):139–45. http://dx.doi.org/10.1016/0167-4048(88)90327-6.

[18] Kephart JO, White SR. Directed-graph epidemiological models of computer viruses. In: Proceedings. 1991 IEEE Computer Society Symposium on Research in Security and Privacy. IEEE; 1991, p. 20–2.

[19] Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. Phys Rev Lett 2001;86(14):3200–3. http://dx.doi.org/10.1103/PhysRevLett.86.3200.

[20] Mahboubi A, Camtepe S, Ansari K. Stochastic Modeling of IoT Botnet Spread: A Short Survey on Mobile Malware Spread Modeling. IEEE Access 2020;8:228818–30. http://dx.doi.org/10.1109/ACCESS.2020.3044277.

[21] Levy N, Rubin A, Yom-Tov E. Modeling infection methods of computer malware in the presence of vaccinations using epidemiological models: an analysis of real-world data. Int J Data Sci Anal 2020;10(4):349–58. http://dx.doi.org/10.1007/s41060-020-00225-1.

[22] Xia H, Li L, Cheng X, Cheng X, Qiu T. Modeling and analysis botnet propagation in social Internet of Things. IEEE IoT J 2020;7(8):7470–81. http://dx.doi.org/10.1109/JIOT.2020.2984662.

[23] ElSawy H, Kishk MA, Alouini M-S. Spatial firewalls: Quarantining malware epidemics in large-scale massive wireless networks. IEEE Commun Mag 2020;58(9):32–8. http://dx.doi.org/10.1109/MCOM.001.2000062.

[24] Nwokoye CH, Madhusudanan V. Epidemic models of malicious-code propagation and control in wireless sensor networks: An indepth review. Wirel Pers Commun 2022;125(2):1827–56. http://dx.doi.org/10.1007/s11277-022-09636-8.

[25] Chernikova A, Gozzi N, Boboila S, Angadi P, Loughner J, Wilden M, et al. Cyber network resilience against self-propagating malware attacks. In: Computer security. Cham, Switzerland: Springer; 2022, p. 531–50. http://dx.doi.org/10.1007/978-3-031-17140-6_26.

[26] Carnier RM, Li Y, Fujimoto Y, Shikata J. Exact Markov Chain of Random Propagation of Malware With Network-Level Mitigation. IEEE IoT J 2023;10(12):10933–47. http://dx.doi.org/10.1109/JIOT.2023.3240421.

[27] Wu G, Xie L, Zhang H, Wang J, Shen S, Yu S. STSIR: An individual-group game-based model for disclosing virus spread in Social Internet of Things. J Netw Comput Appl 2023;214:103608. http://dx.doi.org/10.1016/j.jnca.2023.103608.

[28] Chernikova A, Gozzi N, Perra N, Boboila S, Eliassi-Rad T, Oprea A. Modeling self-propagating malware with epidemiological models. Appl Netw Sci 2023;8(1):1–43. http://dx.doi.org/10.1007/s41109-023-00578-z.

[29] Miller JC. Cocirculation of infectious diseases on networks. Phys Rev E 2013;87(6):060801. http://dx.doi.org/10.1103/PhysRevE.87.060801.

[30] Sanz J, Xia C-Y, Meloni S, Moreno Y. Dynamics of interacting diseases. Phys Rev X 2014;4(4):041005. http://dx.doi.org/10.1103/PhysRevX.4.041005.

[31] Granell C, Gómez S, Arenas A. Competing spreading processes on multiplex networks: Awareness and epidemics. Phys Rev E 2014;90(1):012808. http://dx.doi.org/10.1103/PhysRevE.90.012808.

[32] Steinegger B, Iacopini I, Teixeira AS, Bracci A, Casanova-Ferrer P, Antonioni A, et al. Non-selective distribution of infectious disease prevention may outperform risk-based targeting. Nat Commun 2022;13(3028):1–9. http://dx.doi.org/10.1038/s41467-022-30639-3.

[33] Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S. Catastrophic cascade of failures in interdependent networks. Nature 2010;464:1025–8. http://dx.doi.org/10.1038/nature08932.

[34] Ji P, Ye J, Mu Y, Lin W, Tian Y, Hens C, et al. Signal propagation in complex networks. Phys Rep 2023;1017:1–96. http://dx.doi.org/10.1016/j.physrep.2023.03.005.

[35] Artime O, Grassia M, De Domenico M, Gleeson JP, Makse HA, Mangioni G, et al. Robustness and resilience of complex networks. Nat Rev Phys 2024;6:114–31. http://dx.doi.org/10.1038/s42254-023-00676-y.

[36] Helbing D. Globally networked risks and how to respond. Nature 2013;497:51–9. http://dx.doi.org/10.1038/nature12047.

[37] Diekmann O, Heesterbeek JAP, Metz JAJ. On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. J Math Biol 1990;28(4):365–82. http://dx.doi.org/10.1007/BF00178324.

[38] Tom Britton EP, editor. Stochastic epidemic models with inference. Cham, Switzerland: Springer International Publishing; 2019, URL https://link.springer.com/book/10.1007/978-3-030-30900-8.

[39] Ma J. Estimating epidemic exponential growth rate and basic reproduction number. Infect Dis Model 2020;5:129–41. http://dx.doi.org/10.1016/j.idm.2019.12.009.

[40] Aleta A, de Arruda GF, Moreno Y. Data-driven contact structures: From homogeneous mixing to multilayer networks. PLoS Comput Biol 2020;16(7):e1008035. http://dx.doi.org/10.1371/journal.pcbi.1008035.

[41] Dou J, Xie G, Tian Z, Cui L, Yu S. Modeling and analyzing the spatial–temporal propagation of malware in mobile wearable IoT networks. IEEE IoT J 2023;11(2):2438–52. http://dx.doi.org/10.1109/JIOT.2023.3295016.

[42] Hagberg A, Swart PJ, Schult DA. Exploring network structure, dynamics, and function using networkx. In: Proceedings of the 7th python in science conference. 2018, URL https://www.osti.gov/biblio/960616.

[43] Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem 1977;81(25):2340–61. http://dx.doi.org/10.1021/j100540a008.

[44] Fennell PG, Melnik S, Gleeson JP. Limitations of discrete-time approaches to continuous-time contagion dynamics. Phys Rev E 2016;94(5):052125. http://dx.doi.org/10.1103/PhysRevE.94.052125.

[45] Miller JC, Ting T. EoN (Epidemics on Networks): a fast, flexible Python package for simulation, analytic approximation, and analysis of epidemics on networks. J Open Source Softw 2019;4(44):1731. http://dx.doi.org/10.21105/joss.01731.

[46] Molesky MJ, Cameron EA. Internet of Things: An analysis and proposal of white worm technology. In: 2019 IEEE International Conference on Consumer Electronics. IEEE; 2019, p. 11–3.

[47] Diekmann O, Heesterbeek JAP, Roberts MG. The construction of next-generation matrices for compartmental epidemic models. J R. Soc Interface 2010;7(47):873–85. http://dx.doi.org/10.1098/rsif.2009.0386.