Beyond Herd Immunity Against Strategic Attackers

Vilc Rufino, Daniel Menasché, Cabral Lima, Ítalo Cunha, Leandro P de Aguiar, Eitan Altman, Rachid El-Azouzi, Francesco de Pellegrini, Alberto Avritzer, Michael Grottke

Beyond Herd Immunity Against Strategic Attackers

VILC RUFINO, Federal University of Rio de Janeiro and Brazilian Navy, Brazil
DANIEL MENASCHÉ and CABRAL LIMA, Federal University of Rio de Janeiro, Brazil
LEANDRO P. DE AGUIAR, Siemens Corporate Technology, USA
ÍTALO CUNHA, Federal University of Minas Gerais, Brazil
EITAN ALTMAN, INRIA Sophia Antipolis, France
RACHID EL-AZOUZI and FRANCESCO DE PELLEGRINI, University of Avignon, France
ALBERTO AVRITZER, Esulab Solutions, USA
MICHAEL GROTTKE, Friedrich-Alexander-Universität and GfK SE, Germany

Herd immunity, one of the most fundamental concepts in network epidemics, occurs when a large fraction of the population of devices is immune against a virus or malware. The few individuals who have not taken countermeasures against the threat are assumed to have very low chances of infection, as they are indirectly protected by the rest of the devices in the network. Although very fundamental, herd immunity does not account for strategic attackers scanning the network for vulnerable nodes. In face of such attackers, nodes who linger vulnerable in the network become easy targets, compromising cybersecurity. In this paper, we propose an analytical model which allows us to capture the impact of countermeasures against attackers when both endogenous as well as exogenous infections coexist. Using the proposed model, we show that a diverse set of potential attacks produces non-trivial equilibria, some of which go counter to herd immunity; e.g., our model suggests that nodes should adopt countermeasures even when the remainder of the nodes has already decided to do so.

CCS Concepts: • **Security and privacy → Artificial immune systems**.

Additional Key Words and Phrases: Herd immunity, epidemics, analytical models, counter-measures, game theory

1 INTRODUCTION

Malicious software, such as viruses, Internet worms, adware, spyware and botnets [70], continuously threatens the Internet stability posing a wide variety of challenges to system administrators and users. Viral models for the diffusion of malicious software have been part of the mainstream research in network security to model the diffusion of computer worms [6, 23, 40, 78, 79]. Such models are very convenient also to capture the construction of large distributed attack networks known as *botnets* [43, 59], which are pivotal for the emerging paradigm of cybercriminality as a service. In fact, botnets are built in a silent way using epidemic malware diffusion to compromise millions of terminals by malicious codes (malware) and later on perform actions without the knowledge of the legitimated owners. In the last decade, botnets have been leased as support infrastructure in order to perform various types of criminal activities including, e.g., *Distributed Denial of Service* (DDoS) [52] attack campaigns or ramsomware attacks, just to mention the most spectacular ones. Typically the attacker, also called the botmaster, takes control of devices either compromising them

Authors' addresses: Vilc Rufino, vilc.rufino@marinha.mil.br, Federal University of Rio de Janeiro and Brazilian Navy, Rio de Janeiro, RJ, 21941-590, Brazil; Daniel Menasché, sadoc@dcc.ufrj.br; Cabral Lima, cabrallima@ufrj.br, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, 21941-590, Brazil; Leandro P. de Aguiar, leandro.pfleger@siemens.com, Siemens Corporate Technology, 1 Thørväld Circle, Princeton, NJ, USA; Ítalo Cunha, cunha@dcc.ufmg.br, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil; Eitan Altman, eitan.altman@inria.fr, INRIA Sophia Antipolis, Avignon, France; Rachid El-Azouzi; Francesco de Pellegrini, rachid.elazouzi@univ-avignon.fr, francesco.de-pellegrini@univ-avignon.fr, University of Avignon, Avignon, France; Alberto Avritzer, beto@esulabsolutions.com, Esulab Solutions, Plainsboro, NJ, USA; Michael Grottke, michael.grottke@fau.de, Friedrich-Alexander-Universität and GfK SE, Erlangen-Nürnberg, Germany.

using endogenous infections, i.e., from the neighbors in a local network, or using exogenous infections operated from a remote network. Recently, botnets leverage a lethal combination of social engineering and software vulnerabilities, where the infection of local machines is usually performed using viral phishing attacks able to hijack a large base of social network accounts [27, 59] and by leveraging trojans in order to take control of local machines.

Models for virus propagation have been thoroughly studied since the seminal work of Kermack and McKendrick [17, 38], mainly focusing on epidemic thresholds and immunization policies. In the last 20 years, new research lines in computational epidemiology have unveiled the crucial role of network topology in the propagation of epidemics [14, 48]. As a consequence, in order to manage network security, significant effort has been devoted to understand how computer viruses spread in a network, and how to efficiently design countermeasures able to mitigate such threats [26, 30, 68]. Traditionally, many countermeasures account for *herd immunity*, a form of indirect protection of network nodes from infections that occurs when a large percentage of devices becomes immune to an infection, providing a measure of protection for individuals who are not immune [4, 10, 21, 35].

In fact, vaccination is one of the most prevalent countermeasures against the spread of epidemics, since it reduces the fraction of vulnerable nodes [31, 69]. In the realm of computer systems, however, there are light and heavyweight forms of vaccination. Lightweight vaccination is typically performed through the update of anti-virus software. Such updates are executed regularly, usually once a day or once a week, giving rise to the so-called Internet security “cat and mouse game". Actually, as soon as an anti-virus software update is released, by using dedicated bot update modules [7, 8, 32, 64] botmasters change the signature code of the virus and its behavior. Novel fully undetectable versions of the virus are produced and the virus ultimately evolves through multiple generations [1, 9, 46].

New releases of anti-virus software need to cope with such virus evolution, also known as *polymorphism*. In this context, networked nodes are typically modeled using a susceptible-infectious-susceptible (SIS) model [67], according to which they switch over time between being susceptible (S) to the malware infection and being actually infected (I), and then susceptible again after malware removal. Ultimately, most anti-virus products are subscription-based and deploy regular updates to anti-virus databases.

Alternative countermeasures against viruses include very stringent treatments, such as quarantine, e.g., the disconnection of nodes from the network, clean-state restarts with full operating system and firmware upgrades, or the execution of heavyweight anti-virus software [49, 58]. The latter may detect viruses more promptly compared to their lightweight counterparts, at the expense of more significant CPU and memory overhead. For all practical purposes, devices implementing such countermeasures can be assumed to be immune to the target malware.

Among the challenges faced by system administrators, we focus on the dilemma involved in applying stringent countermeasures, whose applicability is often limited by practical considerations. In fact, although countermeasures like vaccination or patching are very effective, they typically cause collateral effects, such as system downtime or slowdown. In some cases, e.g., in *industrial control systems* [65], the resulting performance losses are unacceptable at the business level. Therefore, one needs to trade off the benefits of applying such countermeasures against their corresponding costs, given the probability of infection in the presence of strategic attackers [10, 30].

A further major challenge in network security is the typically autonomous nature of decision making. Given that devices are interconnected, if the owner of a device or a group of devices is not willing to pay for stringent countermeasures and thus decides to take the risks of contamination, she directly impact its neighbors and indirectly impacts other nodes [11, 28, 66]. Hence, *decision makers face a game in which the countermeasure strategy selected by a given user impacts the security landscape of the population as a whole.*

Cost–benefit analyses of vaccination programs usually account for the *positive externality* of vaccination [3, 10]; i.e., in a population where only a few individuals are not immune, these individuals benefit from the vaccination that the others have undergone. Hence, they have less incentive to incur the relative costs of vaccination. Indeed, their rational decision is to *avoid the crowd*, and ignore the vaccine. Such analyses, however, do not account for exogenous infections caused by malicious and strategic attackers.

Nowadays, it is possible to scan the whole IPv4 space in less than an hour, and efficiently detect a few vulnerable nodes [19, 51]. We refer to attackers performing such port scans to find vulnerable nodes as *strategic attackers*, as they can strategically invest their attack budget towards vulnerable users. Vulnerable users, in turn, must *follow the crowd*, i.e., apply a countermeasure although most of the other users have already done so.

In this paper, we consider the problem of determining whether to invest in heavyweight forms of protection accounting for positive and negative externalities of vaccination. Our goals are:

(1) To compute the node infection probability in a network as a function of the rates of endogenous and exogenous infection; i.e., we assess the risks of not applying a stringent countermeasure.
(2) To determine the system equilibria; i.e., given the relative vaccination costs and an estimate of the infection probability, we determine the expected number of agents that incur the heavyweight relative vaccination costs.

To this aim, we propose a simple epidemic model, which extends the multiplicative SIS model and is amenable to steady-state closed-form solutions. We assume an attacker with a limited average infection budget of Λ infections per time unit. Such power is uniformly distributed among N nodes that the attacker identifies as vulnerable. Then, each of such nodes is subject to exogenous infections which occur at rate Λ/N. Such exogenous infections due to strategic attackers limited by a budget, investigated in this work, give rise to a rich set of novel insights in the realm of epidemic models.

We summarize our key contributions as follows.

(1) **Analytical model**: We propose an analytical model which captures positive and negative externalities associated with countermeasures in security games. It accounts for an attacker with a finite budget, leading to a threat model wherein the exogenous infection rate per node decreases as the number of vulnerable nodes grows. The model is simple and tractable, while still having expressive power to capture the trade-offs related to the vaccination of networked nodes (Sections 3-5);
(2) **Infection probability assessment**: We provide simple closed-form expressions to approximate the infection probability as estimated by the proposed model. In particular, one of the proposed approximations is based on Newton's Approximation Method (NAM). The accuracy of the approximation can be arbitrarily increased at the expense of additional computational cost (Section 6 and Appendix A);
(3) **Vaccination game and analysis of equilibria**: We pose a vaccination game in which each player selects a countermeasure as a function of the estimated infection probability. We investigate system equilibria, indicating two extreme regimes; under the first (second) one, the infection probability monotonically increases (decreases) as a function of the size of the vulnerable population, corresponding to a follow-the-crowd (avoid-the-crowd) behavior (Section 7.1);
(4) **Simulations**: We perform experiments using a detailed malware simulator inspired by *Mirai botnet* epidemics under different configuration scenarios. We verify that the proposed model qualitatively captures the simulated botnet behavior (Section 8 and Appendix G).

This paper is organized as follows. Section 2 presents related work. The considered system is briefly introduced in Section 3. Section 4 defines the vaccination game and the concepts of *follow and avoid the crowd* in the presence of a strategic attacker in an epidemic context. After the proposed model has been described in Section 5, Section 6 develops an approximate solution to the model, in closed form. The system equilibria are analyzed in Section 7. Section 8 illustrates some properties of the considered system through simulation experiments and contrast them against our findings obtained using the model, Section 9 presents additional discussion on broader implications of the work and Section 10 concludes the paper.

2 RELATED WORK

There is a vast literature on epidemic models, accounting for transient and stationary aspects [36] as well as endogenous and exogenous infections [2, 74, 76, 77].

In this work, we assume that a transition from an infected to a susceptible state occurs at nodes deploying lightweight countermeasures. Those transitions reflect that an infected node, after lightweight countermeasures, becomes susceptible again for new variants of the same malware. The use of the SIS model to capture those transitions is standard in the literature of epidemic models applied to computer systems.

The classical SIS epidemic model is borrowed from Biology. As such, it captures the propagation of non-intentional viruses. Propagation of malware in a computer network, in contrast, must capture intentional and targeted infections, as pointed out in [5, 22, 43].

One way to capture strategic behavior is to extend the model by *exogenous strategic infections*. Exogenous infections have previously been considered in the realm of biological networks [2, 62, 76]. However, to the best of our knowledge there is no prior epidemic model using exogenous infections to account for strategic attackers with a finite attack budget. In particular, attackers that can scan the whole network in a few hours have been considered by the security community from a systems-oriented standpoint [19, 45, 54] but not from an epidemics point of view. One of our goals is to bridge this gap. *To that aim, we consider exogenous infections per node whose rates depend on the population size. We are unaware of previous works wherein such threat model has been considered (see Section 3.3).*

Network externalities play an important role in the adoption of software and countermeasures. A community of users of a particular software, for instance, benefits from additional members [3, 16], e.g., when accounting for interoperability or collaboration functionality. In [25], network externalities play an important role for strategic decisions taken by each community in an attacker-susceptible environment. In our work, we assume that increasing the number of vulnerable nodes implies decreasing the probability of an exogenous infection towards a tagged, randomly-chosen node.

Maille et al. [42] have also studied network externalities related to security countermeasures, but without accounting for epidemic aspects. Their focus is on financial and economic motivations behind malicious actions, assuming that the number of vulnerable devices is directly proportional to the incentives an attacker has to produce an exploit for that vulnerability. A similar economic perspective from the standpoint of attackers has been considered in [1]. In our research, in contrast, the focus is mainly on strategic attackers who leverage existing exploits, and are able to identify targets by scanning the IP address space.

We indicate that the proposed model gives rise to both stable and unstable equilibria. Those equilibria are similar in spirit to the ones obtained in the analysis of medium access protocols, such as Aloha [33, 34]. Nonetheless, our analysis intrinsically accounts for strategic decision makers, whereas traditional performance models, such as those used to analyze Aloha [33, 34], account for non-strategic agents.

3 SYSTEM DESCRIPTION

3.1 Terminology

Next, we briefly introduce the terminology considered throughout this work.

- A network comprises *nodes* (or users).
- *Heavyweight forms of vaccination*, also referred to as stringent countermeasures, include quarantine or the execution of heavyweight anti-virus software. Devices implementing such countermeasures are assumed to be immune to the target malware.
- *Vaccinated* nodes are those nodes that have applied stringent forms of vaccination (see Figure 1).
- *Lightweight forms of vaccination* are typically performed through the update of anti-virus software. Such updates are executed regularly, giving rise to the so-called Internet security "cat and mouse game" which motivates the SIS model considered in this paper.
- *Vulnerable* nodes are those nodes that are not vaccinated, who implement lightweight forms of vaccination, and are thus subject to infection and recovery. A vulnerable node is either susceptible or infected.
- *Susceptible* users are prone to infection. Once infected, they apply lightweight countermeasures, which cause them to transition back to the susceptible state.
- The *attack budget* of an attacker is the rate of infections per time unit that the attacker can issue.
- *Endogenous* infections are caused by local neighbors. *Exogenous* infections are caused by an attacker whose attack budget is limited.
- The *vaccination cost* refers to aspects such as expenses, downtime, performance overhead, increased system response time, or degraded functionality due to the application of a vaccine.
- The *infection probability* is the expected fraction of time during which a vulnerable node is in the infected state. The infection probability depends on the infection rate and on the curing rate. When deciding which countermeasure to take, a node trades off the relative vaccination cost against the probability of infection.

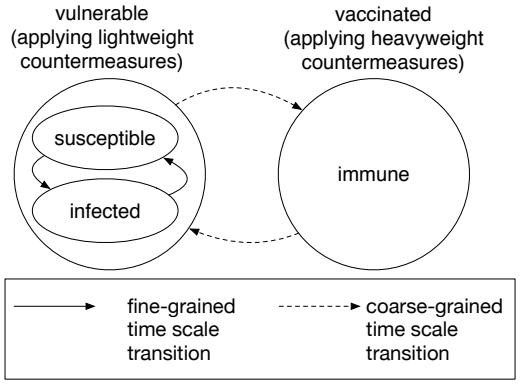


Fig. 1. Node states. Transitions in a coarse-grained time scale correspond to changes in the adoption of stringent countermeasures, and are captured by the vaccination game state graph introduced in Section 4.3 and Definition 4.1. Fine-grained transitions are captured by the SIS model introduced in Section 5.1.

Table 1. Table of notation.

Variable	Description
M	size of population
N	size of vulnerable population (nodes not vaccinated)
γ	endogenous infection rate (per infected edge)
λ	exogenous infection rate (per infected node)
Λ	exogenous infection rate
μ	recovery factor
V	cost per time unit of a vaccine
H	cost per time unit while infected
$C = V/H$	relative vaccination cost
$\boldsymbol{A}$	adjacency matrix
d	number of infected neighboring nodes
$\lambda\gamma^d$	infection rate (per node)
$\pi(\boldsymbol{x})$	probability of state $\boldsymbol{x}$
ι	number of infected nodes
ρ	infection probability (probability of a randomly chosen vulnerable node being infected)
$\hat{\rho}$	infection probability approximated through the binomial approximation
$N^\star$	hyperparameter of binomial approximation applied to fully connected networks (accurate approximations when $N^\star = (N-1)\rho(N)$)
$\bar{d}$	average node degree
$\tilde{d}$	node degree in regular networks where all nodes have same degree ($\tilde{d} = N - 1$ in fully connected networks)
$d^\star$	hyperparameter of binomial approximation, applied when all nodes have roughly the same degree (accurate approximations when $d^\star = \bar{d}\rho(N)$)

3.2 User population, cost function and network topology

We consider a finite population of M nodes. Each of them must decide to invest or not in a vaccine. Let the cost per time unit of a vaccine (e.g., subscription fee for a heavyweight anti-virus software) be denoted by V, while the costs per time unit in which a node is infected amount to H [30]; typically, $V < H$. We refer to the (unit-less) fraction V/H as the *relative vaccination cost* C (see Table 1).

We denote the number of nodes that have not applied a stringent form of vaccination by N. Such nodes are subject to the epidemic process and might be infected by the virus, while still adopting lightweight protective mechanisms. After being infected, a node recovers returning to its initial susceptible state, e.g., by formatting or rebooting the machine. The remaining $M - N$ nodes are assumed to be always immune.

The relative cost incurred by a user in a population wherein N users are vulnerable is given by $C(N)$,

$$C(N) = \begin{cases} \rho, & \text{if user is not vaccinated,} \\ V/H, & \text{otherwise,} \end{cases} \tag{1}$$

where ρ is the fraction of time at which the user is infected.

The network topology, comprising N nodes that did not invest in stringent countermeasures, determines the process of epidemic spread. The network topology is given by its adjacency matrix $\boldsymbol{A}$ of size $N \times N$, where each entry $a_{k,l}$ is 1 if the nodes k and l are connected and 0 otherwise. Except otherwise noted, we assume an *undirected network topology*, wherein $a_{k,l} = a_{l,k}$ and the diagonal elements of $\boldsymbol{A}$ are all zero.

3.3 Threat model

Let Λ be the power of an attacker, measured by the number of infections per time unit. In the simplest setting, a constant budget is allocated evenly among all vulnerable nodes; the *exogenous infection rate per node* is then $\lambda(N) = \Lambda/N$. In the remainder of the paper, we may refer to $\lambda(N)$ simply as λ, keeping the dependence of λ on N implicit but noting that such dependence is assumed throughout the whole work.

In general, Λ may be a function of N, and λ assumes a functional form given by

$$\lambda(N) = \Lambda(N)/N. \tag{2}$$

The threat model introduced above constitutes one of our key contributions. The model leads to novel insights on epidemic behavior accounting for strategic attackers, with implications on the role of vaccination for small populations as further discussed in the following sections.

3.4 Epidemic infection and recovery

At any point in time, each of the N vulnerable nodes can be at states susceptible (S, or 0) or infected (I, or 1). Let the time expended in the recovery of an infected node follow an exponential distribution with rate μ. A susceptible node may be infected by an external attacker (exogenous infection) or by an internal attack (endogenous infection) from network neighbors. Let d be the number of infected neighbors of a given node. We assume that the endogenous infection is exponentially dependent on d; i.e., the rate of endogenous infection per node is given by γ^d. The effect of the exogenous infection is also assumed as multiplicative. Thus, the infection rate of a susceptible node is given by $\lambda\gamma^d$, and the time until a susceptible node becomes infected follows an exponential distribution with mean $1/\left(\lambda\gamma^d\right)$.

3.5 Why multiplicative infection model?

The multiplicative infection model proposed in [75] is inspired by the standard SIS equations. The key novelty consists of replacing the additive infection rate affecting a tagged node, namely $\lambda + \gamma d$, by a multiplicative one $\lambda\gamma^d$. In what follows, we further discuss the motivation and the implications of a multiplicative model.

In the traditional SIS epidemic models inspired by biological systems, such as, the framework under which the NIMFA model [63] is derived, the additive model is a natural choice. In fact, when d infected neighbors of node i enter in contact with node i, each according to an independent Poisson process of rate γ, the resulting cumulative infection rate γd is the sum of their individual infection rates, and yields the Markovian structure. Ultimately, this provides an appealing precise formal derivation of the probability of infection per node.

Note, however, that a linear model may fail to capture the presence of strategic attackers. In fact, such attackers can intentionally target a vulnerable node, possibly in a coordinated and/or often synchronised fashion. As a consequence, our intuition is that such effect might result in a *superlinear infection rate*. Then, a multiplicative model is preferred in that, as showed in [75], it results into *closed-form analytical expressions*, amenable to further analysis under general topologies, as indicated in the upcoming sections.

The additive model [63] captures a situation where the infection rate γd affecting a node increase monotonically as the number of infected neighbors d grows. Under the multiplicative model proposed in [75], in contrast, the cumulative infection rate γ^d may increase or decrease monotonically with respect to the number of infected neighbors d, depending on whether $\gamma \geq 1$ or $\gamma < 1$, respectively. In this paper, we are interested in the scenario wherein the endogenous infection rate increases with respect to the number of infected neighbors. To this aim, we shall assume that the time scale of the epidemic process is rescaled in a way such that the exogenous and endogenous infection rates λ and γ are both greater than or equal to one, $\lambda \geq 1$ and $\gamma \geq 1$.

In Section 8 we show through comprehensive simulations that the qualitative behavior under the multiplicative infection model captured by our analytical results also holds under the additive infection model corresponding to a Mirai botnet. The additive and multiplicative models are further contrasted in Appendix B and analyzed under the complete and bipartite topologies in Appendices D, E and F.

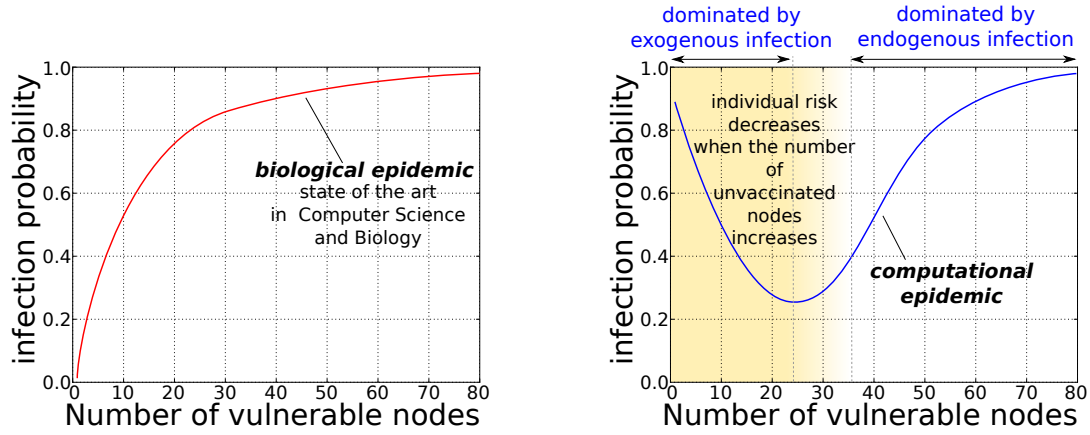
4 THE VACCINATION GAME: FOLLOW THE CROWD OR AVOID IT?

4.1 The two regimes

We distinguish *biological epidemic processes*, which spread infections throughout neighbors without a planned strategy and *computational epidemic processes* which may have two distinct regimes depending on the *i*) the attacker who knows the vulnerable nodes and can directly infect them all subject to its limited capacity or *ii*) the epidemic process that spreads without a direct attacker control (see Figure 2).

In a biological epidemic the infection probability is strictly increasing as a function of the number of vulnerable individuals. This occurs because endogenous infections play a key role and exogenous infections are typically assumed to be insensitive to the number of vulnerable individuals. Such assumptions are captured, for instance, by the standard SIS model under which the NIMFA approximation [63] is derived (Figure 2(a)).

In a computational epidemic considered in this work we assume that exogenous infections are due to a strategic attacker with a finite budget (see Section 3.3). Then, there is an *initial regime* (the yellow area in Figure 2(b)) in which exogenous infection dominates and the infection probability decreases as a function of the number of vulnerable nodes, given that the attacker has limited capacity; and there is a *final regime* in which the endogenous infection dominates and the infection probability increases as a function of the number of vulnerable nodes, similarly to the biological epidemic.



(a) Classical epidemic models [62]

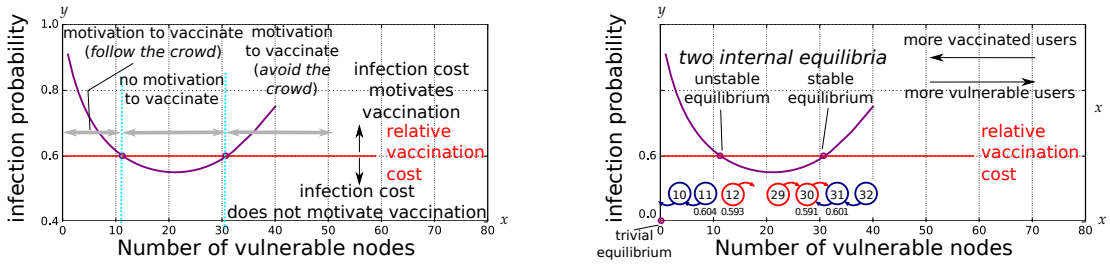
(b) New epidemic model

Fig. 2. Illustrative biological and computational epidemics. Computational epidemic has two types of behavior: the first is dominated by exogenous infections, while the second is dominated by endogenous infections.

4.2 Follow or avoid the crowd?

Figure 3 illustrates in red the relative vaccination costs C for a computational epidemic. If the risk (probability) of infection of a node is above these relative vaccination costs, the node is motivated to vaccinate. Inversely, if the probability of infection of a node is below the relative vaccination costs, the node is not motivated to vaccinate.

According to Figure 3(a), this vaccination strategy leads to three decision moments: *i*) the first moment occurs under the initial regime with the infection probability above the relative vaccination costs: each node is motivated to vaccinate,



(a) Motivation to vaccinate (b) Equilibria

Fig. 3. Infection probability of a tagged vulnerable node: (a) in a computational epidemic, nodes have motivation to vaccinate when the relative vaccination cost is less than the infection probability; (b) the system admits at most two non-trivial equilibria, one being unstable and the other stable.

and hence the number of vulnerable nodes decreases; in this case, the best strategy is to follow the crowd. *ii*) the second moment occurs when the infection probability is below the relative vaccination costs: no node is motivated to vaccinate, and thus the number of vulnerable nodes tends to increase; in this case, the best strategy is to avoid the crowd. *iii*) the third moment occurs under the final regime with the infection probability above the relative vaccination costs: the node is motivated to vaccinate and the number of vulnerable nodes decreases; in this case, the best strategy is to follow the crowd.

4.3 Equilibria

Next, we further explain and formally define the notion of equilibrium considered throughout this paper. We start by illustrating the concepts through an example which is simple but already helps us appreciate the nature of our definitions. Then, we proceed by introducing the formal definitions.

Figure 3(b) shows three points of equilibria: *i*) The first point is the trivial equilibrium, in which there is no vulnerable node and no infected node. *ii*) The second point is an internal unstable equilibrium; few steps towards to the left with respect to the numbers of vulnerable nodes (x axis) implies more motivation to vaccinate, and few steps towards the right implies less motivation to vaccinate. *iii*) The third point is the internal stable equilibrium, a small modification in the number of vulnerable nodes (x axis) to the left or the right results in incentives to return to the equilibrium.

For each vaccination game we define its corresponding state graph. The state graph of the game illustrated above is shown at the bottom of Figure 3(b).

Definition 4.1 (State Graph [24, 41])**.** *A state graph is a directed graph where each vertex corresponds to a strategy profile $\mathcal{Z}$. There is a directed arc from vertex $\mathcal{Z}$ to vertex $\mathcal{Z}'$ with label v if the only difference between $\mathcal{Z}$ and $\mathcal{Z}'$ is the strategy of a single player and the payoff of that player in $\mathcal{Z}$ is strictly less than its payoff in $\mathcal{Z}'$, the modulus of the difference being equal to v.*

Next, we specialize the above general definition of state graphs to the vaccination games considered in this paper. In particular, we consider two simplifying assumptions:

- **symmetry**: we assume all users to be symmetric, i.e., all users have the same number of neighbors and are subject to the same curing rates, as well as the same stationary exogenous and endogenous infection rates. *This*

yields a lumped state space wherein each state is characterized solely by the number of vulnerable users, i.e., the number of users that decided not to implement stringent countermeasures;

- **incentives**: we assume that the infection probability of vulnerable users together with the relative vaccination costs at the current state of the state graph fully determine the incentives that drive users to change their strategies, i.e., *users have an incentive to change their strategy if the current infection probability is greater than relative vaccination costs.*

Intuitively, the latter assumption implies that each user does not account for the difference in the infection probability of the population after a single change of individual strategy is performed. Such assumption allows us to determine the value v of an edge from state $\mathcal{Z}$ to $\mathcal{Z}'$ of the state graph solely based on properties of state $\mathcal{Z}$. Such assumption is inspired by [29], wherein its applicability and implications are further discussed.

Definition 4.2. *The state graph of a vaccination game consists of $N+1$ vertices, with each vertex $n \in \{0, 1, \ldots, N\}$ corresponding to a strategy profile wherein there are n vulnerable users, and each edge corresponding to a transition wherein the system state decreases (or increases) by one unit, representing the fact that a user starts (or stops) adopting a stringent countermeasure. The value v of an edge from state $\mathcal{Z}$ to $\mathcal{Z}'$ is given by*

$$v(\mathcal{Z}) = |H\rho(\mathcal{Z}) - V|, \tag{3}$$

where $\rho(\mathcal{Z})$ is the infection probability at state $\mathcal{Z}$. In addition,

$$\mathcal{Z}' = \begin{cases} \mathcal{Z} - 1, & \text{if } H\rho(\mathcal{Z}) > V \text{ and } \mathcal{Z} \geq 1 \\ & \text{(incentive to start adopting stringent} \\ & \text{countermeasure)} \\ \mathcal{Z} + 1, & \text{if } H\rho(\mathcal{Z}) < V \text{ and } \mathcal{Z} \leq N - 1 \\ & \text{(incentive to stop adopting stringent} \\ & \text{countermeasure),} \\ \mathcal{Z}, & \text{otherwise.} \end{cases} \tag{4}$$

Given the definition of state graphs of vaccination games, we are ready to introduce the notion of stable and unstable equilibria of such games. Note that according to (4), at state $\mathcal{Z}$ a user has *incentive* to adopt (resp., stop adopting) a stringent countermeasure if $H\rho(\mathcal{Z}) > V$ (resp.,$H\rho(\mathcal{Z}) < V$). In what follows, we formalize the notion of an equilibrium.

Definition 4.3. *An equilibrium of the vaccination game is characterized by a minimal set of up to two adjacent vertices n and $n+1$ in its state graph such that there exists a value $n' \in [n, n+1]$ for which $\rho(n') = V/H$, $n' \in \mathbb{R}$, $n \in \mathbb{N}$.*

Definition 4.3 subsumes that the index of each vertex in the state graph corresponds to the expected number of vulnerable users in the system at that state, assuming a large population of users. Under such an interpretation, two adjacent states in the state graph are now separated by a continuum set of virtual states in between them. Then, a virtual equilibrium is a virtual state wherein relative vaccination costs equal infection probability. Accordingly, Definition 4.3 refers to the set of states surrounding that virtual equilibrium state as an equilibrium.

Definition 4.4. *A stable equilibrium of the vaccination game is an equilibrium comprising up to two adjacent vertices n and $n+1$ in its state graph wherein users have no incentive to change their strategies and cause the system to transition to a vertex of the state graph outside of the considered set.*

Note that Definition 4.4 is rather intuitive, as it captures the notion of a set of strategy profiles such that users have no incentive to make the system transition out of this set.

The set of vertices corresponding to a stable equilibrium may comprise a single vertex n or a pair of adjacent vertices n and $n+1$. If an equilibrium comprises two states n and $n+1$, and at state n (state $n+1$) the infection probability is less (greater) than the relative vaccination costs, the equilibrium is stable as (4) implies that the population indefinitely transitions back and forth between those two states.

Definition 4.5. *Any non stable equilibrium of the vaccination game is referred to as an unstable equilibrium.*

According to Definition 4.3, an equilibrium of the vaccination game is characterized by a minimal set of up to two adjacent vertices n and $n+1$ in its state graph. If at state n (state $n+1$) the infection probability is greater (less) than the relative vaccination costs, Definition 4.5 together with (4) imply that the equilibrium is unstable.

Next, we further distinguish between boundary and internal equilibria.

Definition 4.6. *A boundary equilibrium of the vaccination game is an equilibrium corresponding to either vertex 0 or vertex N of the corresponding state graph. Any equilibrium that is not a boundary one is referred to as an internal equilibrium.*

The fact that the infection probability is zero at state 0 and that relative vaccination costs are assumed to be non-negative, together with the two considered simplifying assumptions, motivates the following definition.

Definition 4.7. *The trivial equilibrium is the boundary equilibrium wherein all nodes are vaccinated, corresponding to vertex 0 of the state graph.*

The above definitions will be used in Section 7 to establish structural results of the vaccination game. To that aim, we first introduce the SIS epidemic model and its approximate solution in the two sections that follow. The role of the SIS model and its approximate solution in the general framework considered in this paper is illustrated in Figures 4(a) and 4(b), respectively.

5 EPIDEMIC MODEL: CHARACTERISTICS AND SOLUTION

5.1 Network state

The *network state* can be expressed by an N-dimensional vector. Let $\boldsymbol{x}$ be a state of the network, $\boldsymbol{x} = (x_1, x_2, \ldots, x_k, \ldots, x_{N-1}, x_N)$, with $x_k \in \{0, 1\}$ representing the state of node k and $\boldsymbol{x} \in \mathcal{X}$, with $\mathcal{X} \equiv \{0, 1\}^N$ denoting all possible network states. The dynamics of the system is characterized by a continuous, homogeneous-time, irreducible and finite Markovian process. Each network state corresponds to a state in the Markovian process. Such process, in turn, is known to be reversible [37].

Note that the network states introduced in this section should not be confused with the states of the state graph introduced in Section 4.3 and Definition 4.1. Whereas the states considered here vary in a fine-grained time scale, the states of the state graph considered in Section 4.3 correspond to changes in the adoption of stringent countermeasures by users, and vary in a coarse-grained time scale (see Figure 1).

5.2 Infinitesimal generator

Let $\boldsymbol{Q}$ be the *infinitesimal matrix* associated with the Markov process. States are indexed lexicographically, and we denote by $\boldsymbol{x}^{(i)}$ the *i-th* network state. The *k-th* entry of vector $\boldsymbol{x}^{(i)}$ is denoted by $x_k^{(i)}$. Let $d_k^{(i)}$ be the number of infected neighbors of node k at state $\boldsymbol{x}^{(i)}$.

Then, the element $q_{i,j}$ in the *i-th* row and *j-th* column of $\boldsymbol{Q}$ is given by:

$$q_{i,j} = \begin{cases} \lambda\gamma^{d_k^{(i)}}, & \text{if } x_k^{(i)} = 0, x_k^{(j)} = 1, \\ & x_l^{(i)} = x_l^{(j)} \text{ for } l \neq k, \\ \mu, & \text{if } x_k^{(i)} = 1, x_k^{(j)} = 0, \\ & x_l^{(i)} = x_l^{(j)} \text{ for } l \neq k, \\ -\sum\limits_{p=1,\, p\neq i}^{2^N} q_{i,p}, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

5.3 Steady-state distribution

The steady-state distribution of the multiplicative SIS process [75, 76] is given by

$$\pi(\boldsymbol{x}) \quad = \quad \frac{\tilde{\pi}(\boldsymbol{x})}{Z}, \boldsymbol{x} \in \mathcal{X}, \tag{6}$$

where

$$\tilde{\pi}(\boldsymbol{x}) = \left(\frac{\lambda}{\mu}\right)^{\mathbf{1}^T\boldsymbol{x}} \gamma^{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}/2} \tag{7}$$

and

$$Z = \sum_{\boldsymbol{x}\in\mathcal{X}} \tilde{\pi}(\boldsymbol{x}). \tag{8}$$

The number of infected nodes at state $\boldsymbol{x}$ is given by $\mathbf{1}^T\boldsymbol{x} = \sum\limits_{k=1}^{N} x_k$. In addition, the number of edges with both sides infected, referred to as *infected edges*, is given by $\frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = \frac{1}{2}\sum\limits_{k=1}^{N}\sum\limits_{\substack{l=1\\ l\neq k}}^{N} x_k x_l a_{k,l}$.

5.4 Infection probability

Let I be a random variable denoting the number of infected nodes in the network. Let $\pi(\iota) = P(I = \iota)$ be the probability of finding ι infected nodes in the network. Thus, from Equation (7):

$$\tilde{\pi}(\iota) = \sum_{\boldsymbol{x}:\mathbf{1}^T\boldsymbol{x}=\iota} \tilde{\pi}(\boldsymbol{x}), \quad \iota = 0, \ldots, N \tag{9}$$

$$\pi(\iota) = \frac{\tilde{\pi}(\iota)}{Z}. \tag{10}$$

The infection probability of a node picked randomly (based on a uniform distribution), as a function of the population size, is

$$\rho(N) = \frac{\mathbb{E}(I)}{N}, \tag{11}$$

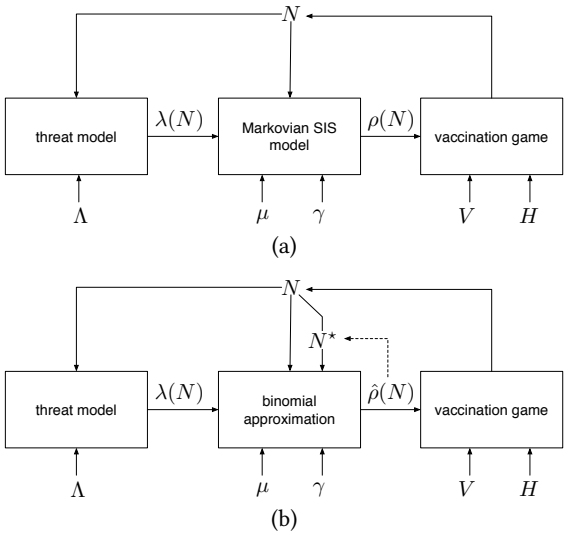


Fig. 4. Exact and approximate solutions to the epidemic model and corresponding vaccination game. The threat model used to obtain the exogenous infection rate $\lambda(N)$ per node is one of the key novel elements of this work.

where

$$\mathbb{E}(I) = \sum_{\iota=0}^{N} \iota \frac{\tilde{\pi}(\iota)}{Z} \tag{12}$$

is the expected number of infected nodes. The infection probability plays a key role in the modeling framework proposed in this work, as summarized in Figure 4(a).

In the expressions above, matrix $\boldsymbol{A}$ is the adjacency matrix as defined in Section 3.2. In the remainder of this paper, we will consider a fully-connected network, unless otherwise noted. For such a network, $a_{k,l} = 1 \; \forall k \neq l$.

6 AN APPROXIMATE SOLUTION TO THE EPIDEMIC MODEL

In this section we introduce an approximate solution to the epidemic model. We start by presenting the binomial approximation.

6.1 Binomial approximation

In what follows, we assume that the topology is fully connected. Let ι be the number of infected nodes. Thus, from Equation (7):

$$\tilde{\pi}(\iota) = \binom{N}{\iota} \left(\frac{\lambda(N)}{\mu}\right)^{\iota} \gamma^{\iota(\iota-1)/2}, \quad \iota = 0, \ldots, N. \tag{13}$$

The infection probability of a node picked randomly (based on a uniform distribution), as a function of the population size, is given by (11),

$$\rho(N) = \frac{1}{N}\sum_{\iota=0}^{N} \iota \frac{\tilde{\pi}(\iota)}{Z} \tag{14}$$

$$= \frac{1}{NZ}\sum_{\iota=0}^{N} \iota \binom{N}{\iota} \left(\frac{\lambda(N)}{\mu}\right)^{\iota} \gamma^{\iota(\iota-1)/2}, \tag{15}$$

where (15) is obtained by replacing (13) into (14). Obtaining a closed-form expression from Equation (15) is complicated due to the quadratic term in the exponent of γ. To simplify this, we consider the following approximation.

Let $N^{\star}(N)$ be an hyperparameter of the proposed approximation of $\rho(N)$. We will show that letting $N^{\star}(N)$ be the expected number of infected neighbors of a typical node yields accurate approximations of $\rho(N)$ in regular topologies wherein all nodes have the same number of neighbors. We defer this derivation to the upcoming section (see also Appendices C and D.5). For now, it suffices to note that $N^{\star}(N)$ is a scalar value between 0 and N and that $N^{\star} : \mathbb{R} \to \mathbb{R}$ is an increasing function.

Then, we define $\hat{\rho}(N) \approx \rho(N)$ and $\hat{\pi}(\iota) \approx \tilde{\pi}(\iota)$ as a function of $N^{\star}(N)$ and of parameters λ, μ, γ and N as follows:

$$\hat{\rho}(N) = \frac{1}{N}\sum_{\iota=0}^{N} \iota \frac{\hat{\pi}(\iota)}{\hat{Z}}, \tag{16}$$

where

$$\hat{\pi}(\iota) = \binom{N}{\iota} \left(\frac{\lambda(N)}{\mu}\gamma^{N^{\star}}\right)^{\iota} \tag{17}$$

and

$$\hat{Z} = \sum_{\iota=0}^{N} \hat{\pi}(\iota). \tag{18}$$

$N^{\star}$ in Equation (17) is a simplified notation for $N^{\star}(N)$. We refer to Equation (16) as a *"binomial approximation"* due to the use of Newton's binomial in its definition.

The role of $N^{\star}$ in the modeling framework proposed in this work is indicated in Figure 4(b), which should be contrasted against Figure 4(a). Given $N^{\star}$, Equation (16) can be rewritten in closed form as demonstrated by Lemma 6.1.

Lemma 6.1. *The node infection probability under the binomial approximation is given by*

$$\hat{\rho}(N) = \frac{1}{1 + \frac{\mu}{\lambda(N)}\gamma^{-N^{\star}}}. \tag{19}$$

Proof. Some algebraic manipulations result in

$$\begin{aligned}\hat{Z}\hat{\rho}(N) &= \frac{1}{N}\sum_{\iota=0}^{N}\iota\binom{N}{\iota}\left(\frac{\lambda(N)}{\mu}\gamma^{N^\star}\right)^{\iota}\\ &= \sum_{\iota=1}^{N}\binom{N-1}{\iota-1}\left(\frac{\lambda(N)}{\mu}\gamma^{N^\star}\right)^{\iota}\\ &= \sum_{\iota=0}^{N}\binom{N-1}{\iota}\left(\frac{\lambda(N)}{\mu}\gamma^{N^\star}\right)^{\iota+1}\\ &= \left(\frac{\lambda(N)}{\mu}\gamma^{N^\star}\right)\left(1+\frac{\lambda(N)}{\mu}\gamma^{N^\star}\right)^{N-1}. \end{aligned} \tag{20}$$

$\hat{Z}$ can be rewritten as

$$\begin{aligned}\hat{Z} &= \sum_{\iota=0}^{N}\hat{\pi}(\iota)\\ &= \sum_{\iota=0}^{N}\binom{N}{\iota}\left(\frac{\lambda(N)}{\mu}\gamma^{N^\star}\right)^{\iota}\\ &= \left(1+\frac{\lambda(N)}{\mu}\gamma^{N^\star}\right)^{N}. \end{aligned} \tag{21}$$

Therefore, by Equations (20) and (21):

$$\begin{aligned}\hat{\rho}(N) = \frac{\hat{\pi}(\iota)}{\hat{Z}} &= \frac{\left(\frac{\lambda(N)}{\mu}\gamma^{N^\star}\right)\left(1+\frac{\lambda(N)}{\mu}\gamma^{N^\star}\right)^{N-1}}{\left(1+\frac{\lambda(N)}{\mu}\gamma^{N^\star}\right)^{N}}\\ &= \left(1+\frac{\mu}{\lambda(N)\gamma^{N^\star}}\right)^{-1}. \end{aligned} \tag{22}$$

□

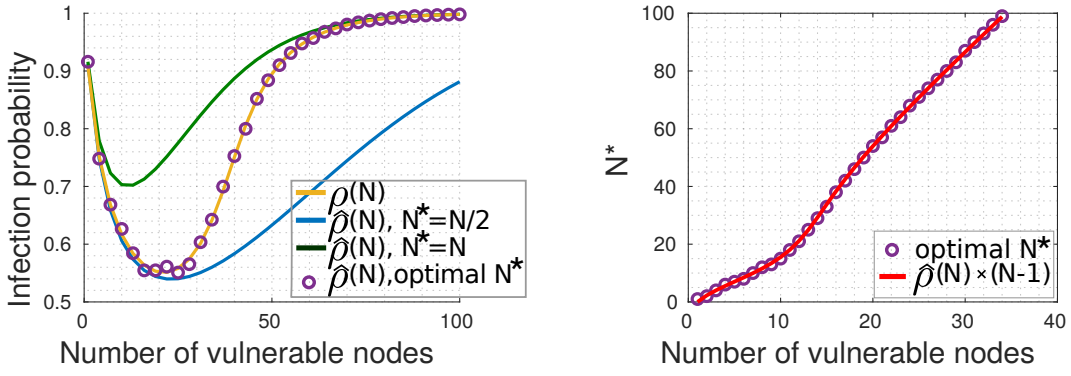
An alternative derivation of the binomial approximation is presented in Appendix C.

6.2 Parameterization of approximation: Optimally setting $N^\star$

Next, our goal is to determine how to set the hyperparameter $N^\star$ in order to obtain accurate results with the proposed approximation of $\rho(N)$. To that aim, we consider two approaches. The first consists of analyzing the most probable states of the system. The second is based on the N-intertwined mean-field approximation (NIMFA). The two approaches are further developed in Appendices C and D.5, respectively. Both approaches lead to the same result, namely, that letting $N^\star$ be the expected number of infected neighbors of a typical node yields accurate approximations in regular networks wherein all nodes have the same degree.

In a fully-connected topology each node has $N-1$ neighbors, and the expected number of infected neighbors of any given node is $(N-1)\rho(N)$. Then, in a fully connected network, under the binomial approximation, Equation (19) can be rewritten as

$$\hat{\rho}(N) = \frac{1}{1+\mu\left(\lambda\gamma^{(N-1)\hat{\rho}(N)}\right)^{-1}}. \tag{23}$$



(a) $\rho(N)$ and its approximations (b) Best $N^\star$ as a function of N

Fig. 5. Infection probability behavior and model parameterization: (a) infection probability $\rho(N)$ as a function of number of vulnerable nodes N, and its approximations $\hat{\rho}(N, N^\star)$, letting $N^\star = N$ (upper bound), $N^\star = N/2$ (lower bound), and optimal $N^\star$; (b) finding the best approximation for $N^\star$ under the binomial approximation, with $\gamma = 1.09$, $\mu = 1$, $\Lambda = 10$ and $\lambda = \Lambda/N$.

where we let

$$N^\star = (N-1)\hat{\rho}(N) \tag{24}$$

in Equation (19).

Figure 5(a) shows the infection probability as a function of the number of vulnerable nodes, with $\gamma = 1.09$, $\mu = 1$, $\Lambda = 10$ and $\lambda = \Lambda/N$. The full orange line is obtained through the fix point solution of Equation (23), which accurately captures the exact solution of the model (see Appendix D). Setting $N^\star$ to its optimal value also leads to an accurate approximation of the infection probability, as indicated by the circles in Figure 5(a). Alternatively, letting $N^\star = N$ (resp., letting $N^\star = N/2$) in Equation (19) leads to an upper (resp., lower) bound for the infection probability, shown by the green (resp., blue) curve.

Figure 5(b) further illustrates how to optimally set $N^\star$. In particular, it indicates that the curves corresponding to the optimal $N^\star$ parameterization and $(N-1)\hat{\rho}(N)$ match each other, which is in agreement with the discussion in the previous paragraphs. The root mean squared error due to the approximation of $N^\star$ by $N\hat{\rho}(N)$ rather than by $(N-1)\hat{\rho}(N)$ is of the order of 10 nodes in the considered setting. This, in turn, indicates that for analytical tractability one may consider simpler expressions to approximate $N^\star$, trading off accuracy against simplicity.

6.3 Regular networks

The analysis in the previous section accounted for fully connected networks, and can be easily extended to regular networks. A regular network is a network wherein each node has degree $\tilde{d}$. In a regular network the expected number of infected neighbors of each node is $\tilde{d}\hat{\rho}(N)$. Then, under the binomial approximation, Equation (19) can be rewritten as

$$\hat{\rho}(N) = \frac{1}{1 + \mu\left(\lambda\gamma^{d^\star}\right)^{-1}}. \tag{25}$$

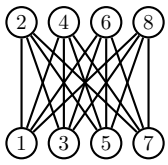
In particular, let

$$d^\star = \tilde{d}\hat{\rho}(N). \tag{26}$$

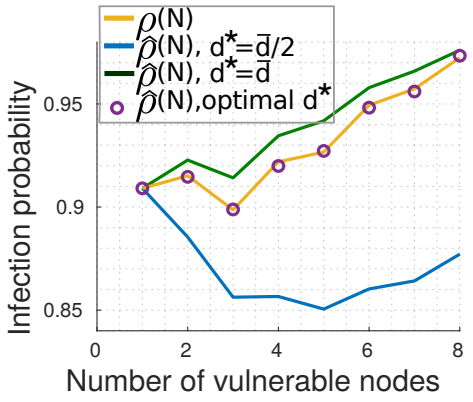
Using (26) we are able to obtain accurate approximations of the infection probability. For the special case of fully connected networks where $\tilde{d} = N - 1$, $\hat{\rho}(N)$ estimated by (25)-(26) equals (23).

Let $\overline{d}$ be the average node degree in a network. If the distribution of node degrees is concentrated around its mean, the analysis above still holds replacing $\tilde{d}$ by $\overline{d}$. In what follows, we illustrate the approximations above in bipartite networks.

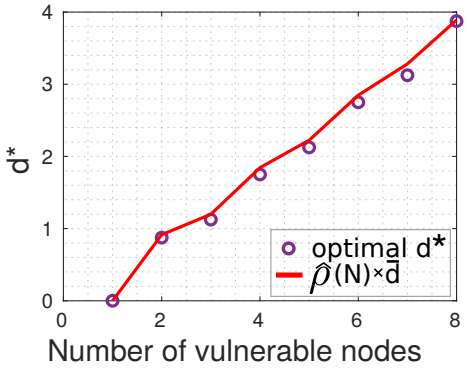
Illustrative example. To illustrate the accuracy of the approximation above, Figures 6 and 7 show the infection probability as a function of the number of vulnerable nodes, obtained through Equation (25). Figure 6 considers a fully connected bipartite network and Figure 7 accounts for a bipartite network with maximum node degree equal 3. In both cases, setting $d^\star$ to its optimal value we obtain very accurate approximations, and (26) typically provides a good approximation for the optimal value of $d^\star$ (see Figures 6(c) and 7(c)). Setting $d^\star = \overline{d}$, we note that the resulting approximation upper bounds the infection probability. Alternatively, setting $d^\star = \overline{d}/2$ we obtain a lower bound. Those examples serve to illustrate that the proposed approximations are helpful to analyze topologies other than the complete graph. Additional results on bipartite graphs and more general topologies are reported in Appendices E and G, respectively.



(a) Fully connected bipartite graph with $N = 8$ nodes

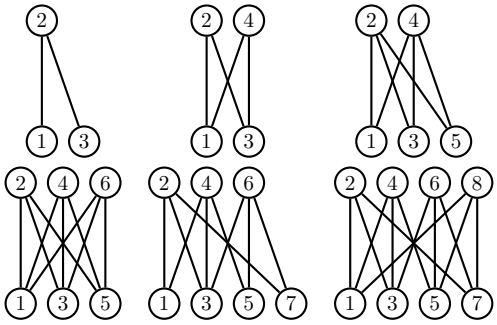


(b) $\rho(N)$ and its approximations in a bipartite graph

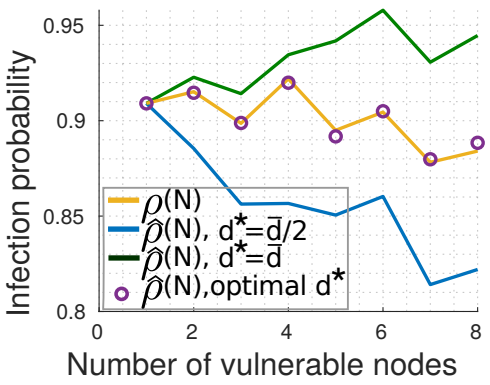


(c) Best $d^\star$ as function of N

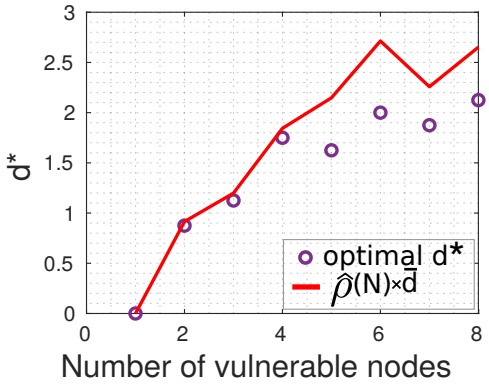
Fig. 6. Infection probability behavior and model parameterization in a bipartite fully connected graph: (a) considered topology when $N = 8$ vulnerable nodes; (b) infection probability $\rho(N)$ and its approximations $\hat{\rho}(N, d^\star)$, letting $d^\star = N/2$ (upper bound), $d^\star = N/4$ (lower bound) and $d^\star$ optimally set; (c) optimal value of $d^\star$ (circles) contrasted against $\rho(N)\overline{d}$ (full line) indicates close agreement between the two. We let $\gamma = 2.09$, $\mu = 1$, $\Lambda = 10$ and $\lambda = \Lambda/N$.



(a) The growth of a bipartite graph with maximum node degree 3, with N varying between 3 and 8



(b) $\rho(N)$ and its approximations in a bipartite graph with maximum node degree 3



(c) Best $d^\star$ as function of N

Fig. 7. Infection probability behavior and model parameterization in a bipartite graph with maximum node degree of three: (a) considered topology when N varies between 3 and 8; (b) infection probability $\rho(N)$ and its approximations $\hat{\rho}(N, d^\star)$, letting $d^\star = N/2$ (upper bound), $d^\star = N/4$ (lower bound) and $d^\star$ optimally set; (c) optimal value of $d^\star$ (circles) contrasted against $\rho(N)\bar{d}$ (full line) indicates rough agreement between the two. We let $\gamma = 2.09$, $\mu = 1$, $\Lambda = 10$ and $\lambda = \Lambda/N$.

7 MODEL ANALYSIS: PROPERTIES OF EQUILIBRIUM

Next, our goal is to characterize structural properties of the equilibria. We start with general results before specializing to the case wherein the attacker budget is distributed uniformly at random across vulnerable nodes.

7.1 General results

Under the general setting illustrated in Figure 4(b), the following theorem states that the model admits at most two internal equilibria, under the mild conditions that $\gamma > 1$ and that $N^\star(N)$ and is an increasing function of N, while $\lambda(N)$ is a decreasing function of N.

Theorem 7.1. *The vaccination dynamics subsumed by the SIS model under the binomial approximation admits at most two internal equilibrium points.*

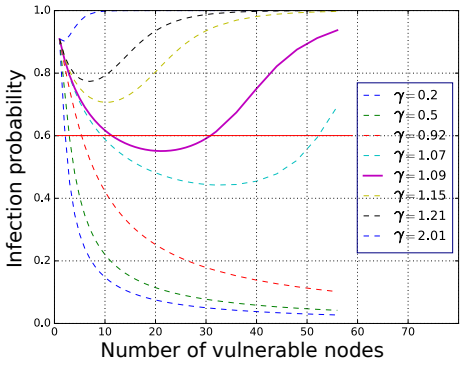


Fig. 8. Infection probability of a tagged node as a function of the population size, $\rho(N)$. When the endogenous infection rate γ is small (big), the system is dominated by the exogenous (endogenous) infection rate, and the infection probability decreases (increases) with respect to the vulnerable population size. When $\gamma \approx 1$, the infection probability first decreases and then increases, in agreement with Equation (6). $\mu = 1$, $\Lambda = 10$ and $\lambda = \Lambda/N$.

Proof. Let $\Psi(N) = (\lambda(N)/\mu)\gamma^{N^\star}$. By Lemma 6.1, $d\hat{\rho}(N)/dN = (d\Psi/dN)/(\Psi^2(1+1/\Psi)^2)$. All terms of $d\hat{\rho}(N)/dN$ are greater than zero, except $d\Psi/dN$:

$$\begin{aligned}\frac{d\Psi}{dN} &= \frac{1}{\mu}\left(\frac{d\lambda}{dN}\gamma^{N^\star} + \lambda(\log\gamma)\gamma^{N^\star}\frac{dN^\star}{dN}\right)\\ &= \frac{\lambda\gamma^{N^\star}}{\mu}\left(\log\gamma\frac{dN^\star}{dN} + \frac{d\lambda}{dN}\frac{1}{\lambda}\right).\end{aligned} \tag{27}$$

As $\frac{dN^\star}{dN} > 0$ and $\frac{d\lambda}{dN} \leq 0$, we conclude that Equation (27) admits at most a single root. Therefore, $\hat{\rho}(N)$ admits at most one internal minimum point, and $\hat{\rho}(N)$ intercepts any horizontal line in at most two points. Those points are the candidate internal equilibria. □

Illustrative example. Figures 8 and 9 illustrate the results discussed so far, under the setup of $\mu = 1$, $\Lambda = 10$ and $\lambda = \Lambda/N$. Theorem 7.1 is in agreement with the results shown in Figure 8. This figure shows that for $\gamma > 1$ the infection probability first decreases and then increases. The infection probability admits a single global minimum and at most two equilibrium points. When the system admits two internal equilibrium points, one of those equilibria is stable, while the other is unstable. For $\gamma = 1.09$, Figure 9 shows the possible population states with corresponding gains envisioned by users who decide to vaccinate (blue arrows pointing upwards) or not to vaccinate (red arrows pointing downwards). States 11 and 12 compose an unstable equilibrium, while states 30 and 31 constitute a stable equilibrium. The minimal infection probability is attained at state 21.

7.2 Special case: vulnerable nodes selected uniformly at random, $\lambda(N) = \Lambda/N$

Next, we specialize our results to the setting wherein vulnerable nodes are selected uniformly at random. To that aim, we leverage the closed-form result derived in Lemma 6.1, setting $\lambda(N) = \Lambda/N$. Note that letting $\lambda(N) = \Lambda/N$ corresponds to considering an attacker who has a finite attack budget of Λ infections per time unit, which is uniformly

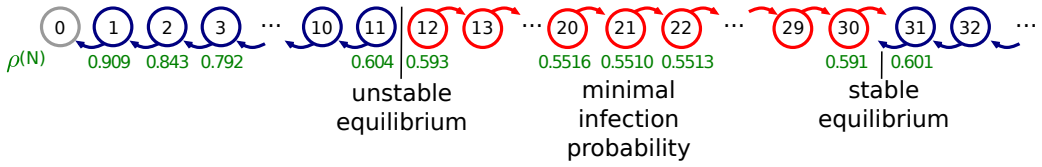


Fig. 9. Dynamics of the number of vulnerable nodes. The endogenous infection rate and the relative vaccination cost are set to $\gamma = 1.09$ and 0.6, using the same setup as for Figure 8. There are two stable Nash equilibria [47], the first one at 0 and the other one at 31. In Figure 8, the horizontal line at $y = 0.6$ crosses the magenta curve at two points. The first point corresponds to an unstable equilibrium (where the expected number of vulnerable nodes is between 11 and 12), and the second one corresponds to a stable equilibrium (between 30 and 31 vulnerable nodes).

distributed across N vulnerable nodes. In that case,

$$\frac{d\lambda}{dN} = -\frac{\Lambda}{N^2}. \tag{28}$$

For the purposes of the following analysis, it suffices to consider a rough approximation for $N^\star$, and let $N^\star = N/2$ (see our discussion in Section 6.2). Then,

$$\frac{d}{dN}\hat{\rho}(N) = \kappa\left(\frac{1}{2}\log\gamma - \frac{1}{N}\right), \tag{29}$$

where κ is a positive constant. The root of Equation (29) corresponds to the population size which yields minimum infection probability, and is given by

$$N = \frac{2}{\log\gamma}. \tag{30}$$

For $\gamma = 1.09$, Equation (30) evaluates to $N \approx 23$. Indeed, this result is in rough agreement with Figure 8, which indicates that the minimum infection probability occurs for $N = 21$.

Next, we further derive closed-form expressions approximating equilibrium points. The internal equilibrium points can be determined through the following equation,

$$\rho(N) - C = 0, \tag{31}$$

where as before C refers to the relative vaccination costs. Alternatively, approximate values can be obtained through

$$\hat{\rho}(N) - C = 0. \tag{32}$$

Letting again $N^\star = N/2$, the values of N that satisfy the above equation are

$$N = -\frac{2}{\log\gamma} W\left(\frac{(C-1)\Lambda\log\gamma}{2C\gamma^{1/2}}\right), \tag{33}$$

where $W(x)$ is the Lambert relation [73], defined as follows,

$$x = W(x)e^{W(x)}. \tag{34}$$

The Lambert relation $W(x)$ admits two real values for each given value of x, corresponding to the branches -1 and 0. For $\Lambda = 10$, $\gamma = 1.09$ and $C = 0.6$, for instance, the values of N corresponding to the -1 and 0 branches are 45.6 and 9.7, respectively. Figure 9 shows that 45.6 significantly overshoots the stable equilibrium involving states 30 and 31, while 9.7 is a good approximation for the unstable equilibrium involving states 11 and 12. The overshooting occurs due to the rough approximation $N^\star = N/2$, which is not very accurate as shown in Figure 5(a). More accurate results can be obtained by using Newton's approximation method (NAM), as indicated in Appendix A.

Although the approximations presented in this section are not extremely accurate, they serve to illustrate the qualitative properties of the model. In particular, the fact that the Lambert relation has at most two real branches implies that the system admits at most two internal equilibria. This result, in turn, is in agreement with Theorem 7.1, allowing us to obtain a quick assessment of the equilibrium points.

8 EXPERIMENTS

We developed an epidemic simulator to evaluate a network's behavior under a wide range of configurations, including those not directly captured in our analytical model (e.g., when the number of vulnerable nodes N varies over time, when the time between infections is not exponentially distributed, or when the epidemic model is additive rather than multiplicative). We compare the experimental results with analytical results, discussing similarities and differences. Our simulator is publicly available.[1]

8.1 Simulator configuration

The simulator provides an array of configuration parameters, shown on Table 2, to allow control of network conditions and the behavior of infected devices. The simulator is inspired by the behavior of Bashlite and Mirai [5, 43].

An attacker operates an initial infection host (called the *master bot*) with an *additive* exogenous infection rate $\tilde{\Lambda}$. On each infection, the master bot attempts connection to a random subset of hosts in the network (e.g., using telnet). Each subsequent infected device (called a *bot*) contributes an *additive* endogenous infection rate $\tilde{\gamma}$. Infections from (master and normal) bots proceed in two steps. First, a bot attempts to connect to a target host (e.g., using telnet). Connections fail if the target device is protected or already infected. Second, the bot attempts to infect the target host if a connection is successful. Each bot attempts infections independently and in a random cyclic order.

We run simulations for $10,000$ time units, which is long enough to estimate the network's steady state. Each configuration was executed eight times; in Figure 10 we plot the infection probability average with a 95% confidence interval as a function of the number of vulnerable hosts.

Table 2. Simulation parameters and their reference values.

Parameter	Description and reference value
Network Properties	
M	Number of nodes; $[10, 28, 46, 64, 82, 100, 200, 300, 400, 500]$.
$N_p = N/M$	Proportion of vulnerable (unvaccinated) nodes; 1.0.
$l_{\min}$, $l_{\max}$	Minimum and maximum round-trip latency, sampled for each pair of nodes from a uniform distribution; 0.01 and 0.4, respectively.
$\mathcal{D}_{\text{on}}$	Distribution of a device's on-time (active) period; exponential with average of 65 time units.
$\tau = \mathbb{E}[\mathcal{D}_{\text{on}}]$	Average host uptime (short notation); 65 time units.
$\mathcal{D}_{\text{off}}$	Distribution of a device off-time (inactive) period. Devices transition to the susceptible state when activated; $\mathcal{D}_{\text{off}}$ is exponentially distributed with average of 0.1 time units.
Malware Behavior	
$\tilde{\Lambda}$	Infection rate from master bot, 2×10^{-2} infections/sec (additive, exogenous infection rate)
$\tilde{\gamma}$	Infection rate from normal bots, 5×10^{-5} inf./sec (additive increase of endogenous infection rate)
T	Connection timeout; time wasted by bots when attempting to connect to protected or infected devices; 2 time units.
m_{conn}	Number of messages in connection attempt, multiplied by the round-trip time between a node pair to determine time taken in the attempt; 7 messages.
m_{infect}	Number of messages in an infection attempt; 700.

[1] https://github.com/queupe/miraisim

8.2 Analytical model and experiments

Model parameterization. Next, we introduce the methodology used to parameterize the proposed model. The most striking distinction between the analytical model and the simulation is that in the former infections have a multiplicative effect, whereas in the latter the effect is additive. For this reason, there is no straightforward mapping between the parameters used in the simulations and those considered in the analytical model. To cope with such a challenge, we consider a simple curve fitting approach. The model parameters corresponding to the scenarios presented in Figure 10 are reported in Table 3. In what follows we further discuss the obtained numerical results.

Table 3. Simulation and fitting of the model parameters

Figure 3	Simulator		Model	
	$\tilde{\gamma}$		μ	γ
(a.I)	8	$\times 10^{-5}$	21.3801	1.0071
(b.I)	20	$\times 10^{-5}$	21.6194	1.0092
(c.I)	50	$\times 10^{-5}$	21.4835	1.0148
(d.I)	500	$\times 10^{-5}$	20.7848	1.0383
	$\tilde{\Lambda}$		μ	γ
(a.II)	5	$\times 10^{-2}$	154.8886	1.0262
(b.II)	32	$\times 10^{-2}$	44.7100	1.0262
(c.II)	200	$\times 10^{-2}$	21.4111	1.0148
(d.II)	2000	$\times 10^{-2}$	15.7172	1.0071
	τ		μ	γ
(a.III)	18	$\times 10^{0}$	122.9879	1.0061
(b.III)	40	$\times 10^{0}$	43.5209	1.0148
(c.III)	65	$\times 10^{0}$	20.9333	1.0148
(d.III)	260	$\times 10^{0}$	2.8819	1.0071

Experimental results. Figure 10 compares the infection probability obtained through simulations against that obtained with the proposed analytical model. The analytical model results were obtained using Newton's Approximation Method after two iterations (see Appendix A). The fraction of infected nodes obtained through simulations (resp., analytical model) is shown in solid red lines (resp., dotted-dashed blue lines). For the simulations, the 95% confidence interval is also reported (shaded area). In addition, we also report the fraction of nodes that were infected through endogenous and exogenous infections, in dotted and dashed red lines, respectively. In each plot, the fraction of infected nodes (solid red line) is the sum of the fraction of endogenously and exogenously infected nodes. Column I (left) varies the endogenous infection rate $\tilde{\gamma}$, Column II (center) varies the exogenous infection rate $\tilde{\Lambda}$, and Column III (right) varies the node uptime τ.

Model validation. The outcome of the experiments is qualitatively in agreement with the findings from the analytical model. Under the initial regime with a few nodes, the system is dominated by exogenous infection. As the number of nodes increases, the infection probability first decreases and then increases, and the system is then dominated by endogenous infection.

Generally, the model tends to overestimate the infection probability vis-à-vis the experiments. This is due to the following reasons: *i*) The model assumes that the nodes are always active (on-time), whereas the simulator assumes that the nodes alternate between active and inactive states (on-time and off-time). *ii*) The model assumes a multiplicative

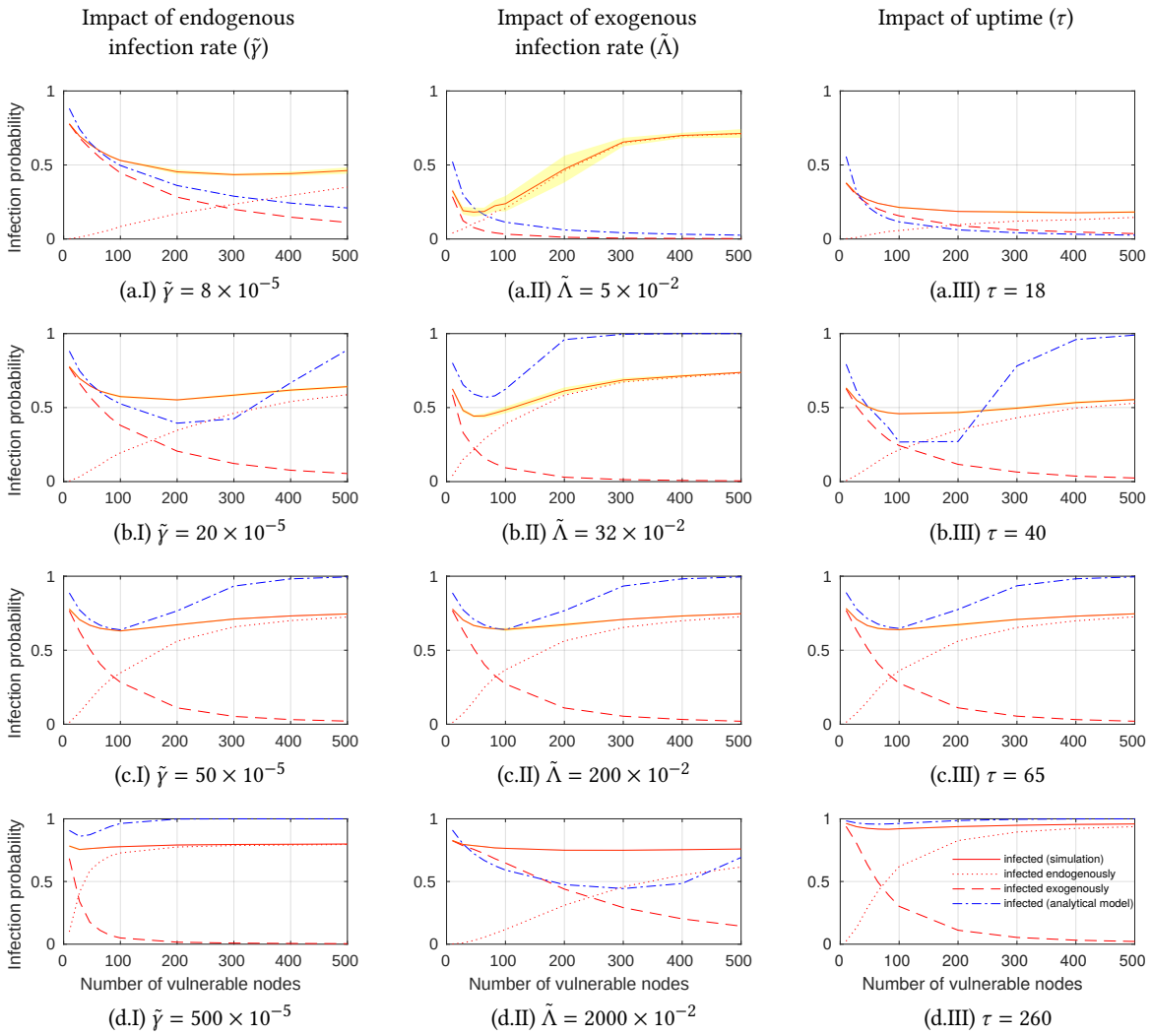


Fig. 10. Outcome of *Mirai Botnet* simulation experiments, in the presence of a strategic attacker, under a fully connected network. The reference values of the simulator parameters are: $\Lambda = 1500$, $\tilde{\gamma} = 5 \times 10^{-5}$, $\tilde{\Lambda} = 2 \times 10^{-2}$ and $\tau = 65$. Model parameters are shown in Table 3.

infection rate, while the simulator assumes an additive one. *iii*) In the model, the periods between events are exponentially distributed, whereas in the simulator l and T follow a uniform distribution. *iv*) The model assumes instantaneous infections and state transitions, while the simulator captures the time it takes bots to scan for vulnerabilities (attempt connections) and infect vulnerable hosts.

8.3 Experimental results and insights

As the number of vulnerable nodes increases, the fraction of infected nodes first decreases and then increases. The system (model and simulator) undergoes two regimes, first being dominated by exogenous infections and

then by endogenous infections. In Figure 10 the dashed curve represents the probability of exogenously infected hosts, which is decreasing while the number of vulnerable hosts increases, and the dotted curve represents the probability of endogenous infected hosts, which is increasing while the number of vulnerable hosts increases. The behavior observed in the experiments agrees with the one predicted by the proposed model (see Lemma 6.1 and Appendix A.2).

Figure 10 shows that under the first regime the proportion of exogenously infected hosts is greater than the proportion of the endogenously infected ones (the dashed curve is above the dotted curve). Under the final regime, the proportion of the endogenously infected hosts is greater than the proportion of the exogenously infected ones (the dotted curve is over the dashed curve).

The minimal proportion of (endogenously and exogenously) infected hosts occurs when the dashed curve crosses the dotted curve (as the curves are concave increasing and convex decreasing, respectively). At that point, the proportion of infected hosts as assessed by simulations (solid curve) reaches its minimum.

Note that when $\tilde{\gamma} = 8 \times 10^{-5}$ (Figure 10(a.I)) the analytical model underestimates the infection probability, going counter the behavior observed in the other scenarios considered in Figure 10. Except for the scenario considered in Figure 10(a.I), the number of vulnerable nodes that minimizes the infection probability according to simulations (solid curve) is typically close to that obtained through the analytical model (dotted-dashed blue curve).

As the number of vulnerable nodes increases, the infection probability is more sensitive to the endogenous, rather than the exogenous, infection rate. Endogenous infections are boosted as the number of infected nodes in the network increases, whereas the exogenous infection rate is limited by the power of the *bot master*. From top to bottom, in Figure 10 Column I the endogenous infection rate was increased by about factor 60, while in Column II the exogenous infection rate had to be increased around 400 times to produce similar effects.

Model parameters are more sensitive to host uptimes and exogenous infection rates as opposed to endogenous rates. As shown in Table 3, model parameters μ and γ are more sensitive to $\tilde{\Lambda}$ and τ (second and third columns of Figure 10) as opposed to $\tilde{\gamma}$ (first column of Figure 10). This occurs even though endogenous infection rate was varied in a range between 8×10^{-5} and 500×10^{-5}, which allows us to appreciate the roles of endogenous and exogenous infections as the number of vulnerable nodes increases. Further increasing the endogenous infection rate does increase the sensitivity of model parameters, however, this results in scenarios where endogenous infection rates dominate system behavior, which can be captured through classical epidemic models, e.g., [17, 38].

Host uptimes significantly impact the fraction of infected nodes. The asymptotic value of the proportion of infected nodes depends on the average uptime, as shown in Figure 10, Column III. When nodes stay active for longer periods of time, the number of infections attempted by each individual bot increases, resulting in an increase of the fraction of infected nodes.

9 DISCUSSION

In this section we indicate some of the broader implications of the results presented in this work.

9.1 Cybersecurity insurance

Cybersecurity insurance (or cyber liability insurance) is a product that an entity can purchase to help reduce the financial risks associated with online business. It encompasses a contract wherein, in exchange for a fee, the insurance policy transfers some of the risk to the insurer [56, 72]. Our results imply that the modeling and pricing of cybersecurity insurance should take into account both positive and negative externalities derived from immunization. In particular, the model proposed in this paper may serve as an additional ingredient when assessing insurance prices [72].

9.2 Risk score parameterization

Standard risk scores, such as the common vulnerability scoring system [44], account for environmental aspects when determining risks. Such environmental aspects may embrace the security countermeasures taken by the neighbors of a node when assessing its risk. Our results indicate that even if most of the neighbors of a given node are already protected, the risks faced by a node may remain high, which serves to motivate lingering nodes to also deploy the available security countermeasures.

9.3 Immunization strategies beyond herd immunity

The models presented in this work serve to bring awareness to system administrators about risks incurred due to old vulnerabilities for which a significant fraction of the population has already applied a patch. Strategic attackers may still be able to find vulnerable nodes that linger in the network. Such nodes may correspond, for instance, to industrial control systems which are difficult to patch, or to devices which are not automatically patched after being installed off-the-shelf [65]. Strategic attackers may target those devices, requiring system administrators to adopt preventive measures beyond herd immunity.

9.4 Additional Practical implications

Next, we provide a discussion on the implications of our results from the attack generation and defense points of view. From the attack standpoint, our model suggests that scanning the network to target vulnerable nodes can significantly impact infection probability. This, in turn, implies that engineering solutions to counteract the automated exploitation of vulnerabilities in the wild are key in face of strategic attackers [18, 19]. From the defense standpoint, our model suggests that there is an optimal number of vulnerable nodes that minimizes infection probability. We envision that the assessment of infection probability, in turn, can be used to decide how to invest in security countermeasures, such as vaccination, rejuvenation and quarantine [26], accounting for the whole population ecosystem. In addition, the number of vulnerable nodes that minimizes infection probability as derived from the proposed model can also be instrumental to determine how to deploy honeypots based on first principles [53], which we leave as subject for future work.

10 CONCLUSION

In this paper we have proposed a new epidemic analytical model to assess the infection probability of nodes in a network which face a strategic attacker with finite power. For this model, the infection probability can be expressed in closed form, allowing us to verify its equilibrium points and its further attributes. Administrators can use this model to choose the best countermeasure to be applied to the network. To facilitate this process, the model provides: *i*) a *vaccination game* in which the player must choose the best strategy to minimize the maintenance costs depending on the infection probability; *ii*) some points of equilibrium supporting the concept of *follow or avoid the crowd* in the presence of a strategic attacker.

In order to validate the proposed model we have carried out numerous experiments using a simulator in which we provoked infections using the *Mirai botnet* malware. The proposed model was able to capture the behavior qualitatively and accurately. The experiments have also allowed to understand what happens if some assumptions of the proposed model are relaxed. The experiments have shown that the exogenous infection rate has to be increased around 400 times to attain effects similar to those observed when the endogenous infection rate is increased by about a factor of 60.

Some interesting results from the experiments and related analysis include the following. *i*) There are two distinct regimes, the first one being dominated by exogenous infections and the second one by endogenous infections. *ii*) The role of endogenous infection is prevalent whenever the number of vulnerable nodes is big. *iii*) In contrast to classical epidemiology research, a few vulnerable nodes may become preferred targets, and increasing the number of vulnerable nodes may decrease the infection probability of a given tagged node. The latter observation, in turn, may be used to position honeypots in a network based on epidemiological first principles, which we leave as subject for future work.

We envision that this work opens up a number of directions for future research, including the analysis of the spread of two or more distinct malware. Protective measures can then be implemented either at the host level, e.g. upgrades to OSes/firmware that add address space layout randomization (ASLR), or at the network level, e.g. blocking SMB ports and protecting against multiple malware using the same attack vector. It is also worth pointing that in this work we focused on networks with a finite number of nodes. The study of scaling laws of epidemics when the number of vulnerable nodes grows to infinity [57], accounting for strategic attackers whose budget increases as the population of vulnerable nodes grows, is another avenue for future research.

REFERENCES

[1] Luca Allodi. 2017. Economic factors of vulnerability trade and exploitation. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 1483–1499.

[2] Eitan Altman, Alberto Avritzer, Rachid El-Azouzi, Daniel S Menasche, and Leandro Pfleger de Aguiar. 2014. Rejuvenation and the spread of epidemics in general topologies. In *2014 IEEE International Symposium on Software Reliability Engineering Workshops*. IEEE, 414–419.

[3] Ross Anderson and Tyler Moore. 2006. The economics of information security. *Science* 314, 5799 (2006), 610–613.

[4] Roy M Anderson and Robert M May. 1985. Vaccination and herd immunity to infectious diseases. *Nature* 318, 6044 (1985), 323.

[5] Manos Antonakakis, Tim April, Michael Bailey, Matt Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J. Alex Halderman, Luca Invernizzi, Michalis Kallitsis, Deepak Kumar, Chaz Lever, Zane Ma, Joshua Mason, Damian Menscher, Chad Seaman, Nick Sullivan, Kurt Thomas, and Yi Zhou. 2017. Understanding the Mirai botnet. In *26th USENIX Security Symposium (USENIX Security 17)*. USENIX Association, Vancouver, BC, 1093–1110.

[6] Alberto Avritzer, Robert G. Cole, and Elaine J. Weyuker. 2010. Methods and opportunities for rejuvenation in aging distributed software systems. *Journal of Systems and Software* 83, 9 (2010), 1568–1578.

[7] Zsolt Bederna and Tamas Szadeczky. 2019. Cyber espionage through Botnets. *Security Journal* (2019), 1–20.

[8] Leon Böck, Emmanouil Vasilomanolakis, Jan Helge Wolf, and Max Mühlhäuser. 2019. Autonomously detecting sensors in fully distributed botnets. *Computers & Security* 83 (2019), 1–13.

[9] Jean-Marie Borello and Ludovic Mé. 2008. Code obfuscation techniques for metamorphic viruses. *Journal in Computer Virology* 4, 3 (2008), 211–220.

[10] Marc Brisson and William J Edmunds. 2003. Economic evaluation of vaccination programs: the impact of herd-immunity. *Medical Decision Making* 23, 1 (2003), 76–82.

[11] Levente Buttyán and Jean-Pierre Hubaux. 2007. *Security and Cooperation in Wireless Networks*. Cambridge University Press.

[12] E Cator, P Donnelly, and P Van Mieghem. 2018. Reply to Comment on Nodal infection in Markovian susceptible-infected-susceptible and susceptible-infected-removed epidemics on networks are non-negatively correlated. *Physical review. E* 98, 2-2 (2018), 026302.

[13] E Cator and P Van Mieghem. 2014. Nodal infection in Markovian SIS and SIR epidemics on networks are non-negatively correlated. *Physical Review E* 89, 5 (2014), 052802.

[14] Deepayan Chakrabarti, Yang Wang, Chenxi Wang, Jurij Leskovec, and Christos Faloutsos. 2008. Epidemic thresholds in real networks. *ACM Transactions on Information and System Security (TISSEC)* 10, 4 (2008), 1.

[15] M. Darboux. 1869. Sur la méthode d'approximation de Newton. In *Nouvelles Annales de Mathématiques*, Vol. 8. 17–27.

[16] Debabrata Dey and Guoying Zhang. 2011. Impact of Network Externality in the Security Software Market. In *Theory in Economics of Information System*.

[17] Odo Diekmann, Johan Andre Peter Heesterbeek, and Johan AJ Metz. 1995. The legacy of Kermack and McKendrick. *Publications of the Newton Institute* 5 (1995), 95–115.

[18] Zakir Durumeric, Frank Li, James Kasten, Johanna Amann, Jethro Beekman, Mathias Payer, Nicolas Weaver, David Adrian, Vern Paxson, Michael Bailey, and J. Alex Halderman. 2014. The Matter of Heartbleed. In *Proceedings of the Internet Measurement Conference (IMC '14)*. 475–488.

[19] Zakir Durumeric, Eric Wustrow, and J Alex Halderman. 2013. ZMap: Fast Internet-wide Scanning and Its Security Applications. In *USENIX Security Symposium*, Vol. 2013.

[20] Chinwendu Enyioha, Ali Jadbabaie, Victor Preciado, and George J Pappas. 2015. Distributed resource allocation for epidemic control. *arXiv preprint arXiv:1501.01701* (2015).

[21] Paul E M Fine. 1993. Herd immunity: history, theory, practice. *Epidemiologic Reviews* 15, 2 (1993), 265–302.

[22] M Todd Gardner, Cory Beard, and Deep Medhi. 2017. Using SEIRS Epidemic Models for IoT Botnets Attacks. In *DRCN 2017-Design of Reliable Communication Networks; 13th International Conference*. 1–8.

[23] Michele Garetto, Weibo Gong, and Don Towsley. 2003. Modeling malware spreading dynamics. In *Proc. of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies*, Vol. 3. 1869–1879.

[24] Michel X Goemans, Li Li, Vahab S Mirrokni, and Marina Thottan. 2006. Market sharing games applied to content distribution in ad hoc networks. *IEEE Journal on Selected Areas in Communications (JSAC)* 24, 5 (2006), 1020–1033.

[25] Jens Grossklags, Nicolas Christin, and John Chuang. 2008. Secure or insure?: a game-theoretic analysis of information security games. In *Proc. 17th International Conference on World Wide Web*. 209–218. DOI: 10.1145/1367497.1367526.

[26] Michael Grottke, Alberto Avritzer, Daniel S Menasché, Leandro P de Aguiar, and Eitan Altman. 2016. On the Efficiency of Sampling and Countermeasures to Critical-Infrastructure-Targeted Malware Campaigns. *ACM SIGMETRICS Performance Evaluation Review* 43, 4 (2016), 33–42.

[27] Surbhi Gupta, Abhishek Singhal, and Akanksha Kapoor. 2016. A literature survey on social engineering attacks: Phishing attack. In *2016 International Conference on Computing, Communication and Automation*. 537–540.

[28] Zhu Han, Dusit Niyato, Walid Saad, Tamer Başar, and Are Hjørungnes. 2012. *Game theory in wireless and communication networks: theory, models, and applications*. Cambridge University Press.

[29] Alain Haurie and Patrice Marcotte. 1985. On the relationship between Nash—Cournot and Wardrop equilibria. *Networks* 15, 3 (1985), 295–308.

[30] Yezekael Hayel, Stojan Trajanovski, Eitan Altman, Huijuan Wang, and Piet Van Mieghem. 2014. Complete game-theoretic characterization of SIS epidemics protection strategies. In *2014 IEEE 53rd Annual Conference on Decision and Control (CDC)*. 1179–1184.

[31] Dirk Helbing, Dirk Brockmann, Thomas Chadefaux, Karsten Donnay, Ulf Blanke, Olivia Woolley Meza, Mehdi Moussaid, Anders Johansson, Jens Krause, Sebastian Schutte, and Matjaž Perc. 2014. Saving Human Lives: What Complexity Science and Information Systems can Contribute. *Journal of Statistical Physics* 158, 3 (Jun 2014), 735–781. https://doi.org/10.1007/s10955-014-1024-9

[32] Azhar Iqbal, Mingyu Guo, Lachlan Gunn, M. Ali Babar, and Derek Abbott. 2019. Game theoretical modelling of network/cyber security. *IEEE Access* 7 (2019), 154167–154179.

[33] Y-C Jenq. 1980. On the stability of slotted ALOHA systems. *IEEE Transactions on Communications* 28, 11 (1980), 1936–1939.

[34] Peter Johansson and Robert Forchheimer. 2019. Information networks slides. https://www.icg.isy.liu.se/courses/tsin01/material/slides/3_MAC2.pdf and https://www.icg.isy.liu.se/courses/tsin01/.

[35] T Jacob John and Reuben Samuel. 2000. Herd immunity and herd effect: new insights and definitions. *European Journal of Epidemiology* 16, 7 (2000), 601–606.

[36] Matt J Keeling and Ken T D Eames. 2005. Networks and epidemic models. *Journal of the Royal Society Interface* 2, 4 (2005), 295–307.

[37] Frank P Kelly. 1979. *Reversibility and Stochastic Networks*. John Wiley, New York.

[38] William Ogilvy Kermack and Anderson G McKendrick. 1927. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London (Series A)* 115, 772 (1927), 700–721.

[39] L.D. Landau and E.M. Lifshitz. 1984. *Statistical Physics (Course of Theoretical Physics, Volume 5)*. Butterworth-Heinemann.

[40] Wanping Liu, Chao Liu, Xiaoyang Liu, Shaoguo Cui, and Xianying Huang. 2016. Modeling the spread of malware with the influence of heterogeneous immunization. *Applied Mathematical Modelling* 40, 4 (2016), 3141–3152. https://doi.org/10.1016/j.apm.2015.09.105

[41] Qian Ma, Edmund Yeh, and Jianwei Huang. 2019. How Bad is Selfish Caching?. In *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*. 11–20.

[42] Patrick Maillé, Peter Reichl, and Bruno Tuffin. 2011. Interplay between Security Providers, Consumers, and Attackers: A Weighted Congestion Game Approach. In *International Conference on Decision and Game Theory for Security*. 67–86.

[43] Artur Marzano, David Alexander, Osvaldo Fonseca, Elverton Fazzion, Cristine Hoepers, Klaus Steding-Jessen, Marcelo HPC Chaves, Ítalo Cunha, Dorgival Guedes, and Wagner Meira. 2018. The Evolution of Bashlite and Mirai IoT Botnets. In *2018 IEEE Symposium on Computers and Communications (ISCC)*. 813–818.

[44] Peter Mell, Karen Scarfone, and Sasha Romanosky. 2006. Common vulnerability scoring system. *IEEE Security & Privacy* 4, 6 (2006), 85–89.

[45] Lionel Metongnon and Ramin Sadre. 2018. Fast and efficient probing of heterogeneous IoT networks. *International Journal of Network Management* 28, 1 (2018).

[46] Philip O'Kane, Sakir Sezer, and Kieran McLaughlin. 2011. Obfuscation: The hidden malware. *IEEE Security & Privacy* 9, 5 (2011), 41–47.

[47] Martin J Osborne and Ariel Rubinstein. 1994. *A Course in Game Theory*. MIT press.

[48] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. 2015. Epidemic processes in complex networks. *Reviews of Modern Physics* 87, 3 (2015), 925.

[49] Nathanael Paul, Sudhanva Gurumurthi, and David Evans. 2005. Towards disk-level malware detection. *Code Based Software Security Assessments* (2005), 13.

[50] Bo Qu and Huijuan Wang. 2016. The Accuracy of Mean-Field Approximation for Susceptible-Infected-Susceptible Epidemic Spreading with Heterogeneous Infection Rates. In *International Workshop on Complex Networks and their Applications*. Springer, 499–510.

[51] Alan Quach, Zhongjie Wang, and Zhiyun Qian. 2017. Investigation of the 2016 Linux TCP Stack Vulnerability at Scale. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 1, 1 (2017), 4.

[52] Anirudh Ramachandran, David Dagon, and Nick Feamster. 2006. Can DNS-based blacklists keep up with bots?. In *CEAS 2016 – The Third Conference on Email and Anti-Spam*. Mountain View, California, USA.
[53] Jianguo Ren and Yonghong Xu. 2018. A compartmental model to explore the interplay between virus epidemics and honeynet potency. *Applied Mathematical Modelling* 59 (2018), 86–99.
[54] Philipp Richter and Arthur Berger. 2019. Scanning the Scanners: Sensing the Internet from a Massively Distributed Network Telescope. In *Proceedings of the ACM Internet Measurement Conference* (Amsterdam, Netherlands). ACM, New York, NY, USA, 144–157.
[55] Pablo M Rodríguez, Alejandro Roldán-Correa, and Leon Alexander Valencia. 2018. Comment on "Nodal infection in Markovian susceptible-infected-susceptible and susceptible-infected-removed epidemics on networks are non-negatively correlated". *Physical Review E* 98, 2 (2018), 026301.
[56] Margaret Rouse. 2019. Cybersecurity insurance. https://whatis.techtarget.com/definition/cybersecurity-insurance.
[57] Sarabjeet Singh and Christopher R Myers. 2014. Outbreak statistics and scaling laws for externally driven epidemics. *Physical Review E* 89, 4 (2014), 042108.
[58] Nina Skorin-Kapov, Marija Furdek, Szilard Zsigmond, and Lena Wosinska. 2016. Physical-layer security in evolving optical networks. *IEEE Communications Magazine* 54, 8 (2016), 110–117.
[59] Kurt Thomas and David M. Nicol. 2010. The Koobface botnet and the rise of social malware. In *5th International Conference on Malicious and Unwanted Software*. 63–70.
[60] Piet Van Mieghem. 2011. The N-intertwined SIS epidemic network model. *Computing* 93, 2-4 (2011), 147–169.
[61] P Van Mieghem. 2016. Approximate formula and bounds for the time-varying susceptible-infected-susceptible prevalence in networks. *Physical Review E* 93, 5 (2016), 052312.
[62] Piet Van Mieghem and Eric Cator. 2012. Epidemics in networks with nodal self-infection and the epidemic threshold. *Physical Review E* 86, 1, Art. no. 016116 (2012).
[63] Piet Van Mieghem, Jasmina Omic, and Robert Kooij. 2009. Virus spread in networks. *IEEE/ACM Transactions on Networking (ToN)* 17, 1 (2009), 1–14.
[64] Benjamin Vignau, Raphael Khoury, and Sylvain Hallé. 2019. 10 Years of IoT Malware: a Feature-Based Taxonomy. In *2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C)*. 458–465.
[65] Brandon Wang, Xiaoye Li, Leandro P de Aguiar, Daniel S Menasche, and Zubair Shafiq. 2017. Characterizing and Modeling Patching Practices of Industrial Control Systems. *Proceedings of the ACM on Measurement and Analysis of Computing Systems (POMACS)* 1, 1 (2017), 18.
[66] Sinong Wang and Ness Shroff. 2017. Security Game with Non-additive Utilities and Multiple Attacker Resources. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 1, 1, Article 13 (2017), 32 pages. http://doi.acm.org/10.1145/3084450
[67] Yang Wang, Deepayan Chakrabarti, Chenxi Wang, and Christos Faloutsos. 2003. Epidemic spreading in real networks: An eigenvalue viewpoint. In *22nd International Symposium on Reliable Distributed Systems*. 25–34.
[68] Zhen Wang, Chris T Bauch, Samit Bhattacharyya, Alberto d'Onofrio, Piero Manfredi, Matjaž Perc, Nicola Perra, Marcel Salathé, and Dawei Zhao. 2016. Statistical physics of vaccination. *Physics Reports* 664 (2016), 1–113.
[69] Zhen Wang, Yamir Moreno, Stefano Boccaletti, and Matjaž Perc. 2017. Vaccination and epidemics in networked populations — an introduction.
[70] Sharon Weinberger. 2011. Computer security: Is this the start of cyberwarfare? *Nature* 474 (2011), 142–145.
[71] Samuel S Wilks. 1932. Moments and distributions of estimates of population parameters from fragmentary samples. *The Annals of Mathematical Statistics* 3, 3 (1932), 163–195.
[72] Maochao Xu and Lei Hua. 2019. Cybersecurity Insurance: Modeling and Pricing. *North American Actuarial Journal* 23, 2 (2019), 220–249.
[73] Sun Yi, Patrick W Nelson, and A Galip Ulsoy. 2010. *Time-delay systems: analysis and control using the Lambert W function*. World Scientific.
[74] June Zhang. 2015. *Network Process: How Topology Impacts the Dynamics of Epidemics and Cascading Failures*. PhD dissertation. Carnegie Mellon University.
[75] June Zhang and José M F Moura. 2014. Diffusion in social networks as SIS epidemics: Beyond full mixing and complete graphs. *IEEE Journal of Selected Topics in Signal Processing* 8, 4 (2014), 537–551.
[76] June Zhang and José M F Moura. 2017. Contact process with exogenous infection and the scaled SIS process. *Journal of Complex Networks* 5, 5 (2017), 712–733.
[77] June Zhang and Jose M. F. Moura. 2018. Who is More at Risk in Heterogenous Networks?. In *Proc. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Calgary, AB, Canada, 1–5.
[78] Cliff Changchun Zou, Lixin Gao, Weibo Gong, and Don Towsley. 2003. Monitoring and early warning for Internet worms. In *Proceedings of the 10th ACM Conference on Computer and Communications Security*. 190–199.
[79] Cliff Changchun Zou, Weibo Gong, and Don Towsley. 2003. Worm propagation modeling and analysis under dynamic quarantine defense. In *Proceedings of the 2003 ACM Workshop on Rapid Malcode*. 51–60.

A ITERATIVE MODEL SOLUTION THROUGH NEWTON APPROXIMATION

In this appendix we indicate that the Newton Approximation Method (NAM) may be instrumental to approximate the infection probability. The use of NAM to obtain approximate closed form expressions to estimators is not novel. A

similar idea has been used by Wilks [71], for instance, to obtain approximations in closed form to optimal estimators for the covariance matrix of the bivariate Gaussian distribution (see Section 4 in [71]).

A.1 A note about notation

In all the appendices that follow, we are interested in approximations to $\rho(N)$. Then, to simplify notation we refer to the infection probability as estimated by the Markovian model and to its approximation through the binomial approximation as $\rho(N)$. It should be clear from the context the quantity which is being referred to.

A.2 Introduction to Newton Approximation Method (NAM)

Next, we show how to apply NAM to obtain approximations for the infection probability. In a fully connected network, under the binomial approximation, Equation (19) can be rewritten as

$$\rho(N) = \frac{1}{1 + \mu \left(\lambda \gamma^{(N-1)\rho(N)}\right)^{-1}}. \tag{35}$$

The definition of $\rho(N)$ uses $\rho(N)$ itself as an exponent of γ. Isolating $\rho(N)$ is non-trivial, but we can approximate it. With $\rho(N) = \rho$ and $n = N - 1$, define function $f(\rho)$ as

$$f(\rho) = \rho \left(1 + \frac{\mu}{\lambda} \gamma^{-\rho n}\right) - 1 = \rho + \rho \frac{\mu}{\lambda} \gamma^{-\rho n} - 1. \tag{36}$$

Then,

$$f'(\rho) \equiv \frac{\partial f(\rho)}{\partial \rho} = 1 + \frac{\mu}{\lambda} \gamma^{-\rho n} \left(1 - \rho n \ln \gamma\right) \tag{37}$$

$$f''(\rho) \equiv \frac{\partial^2 f(\rho)}{\partial^2 \rho} = \frac{\mu n \ln \gamma}{\lambda} \gamma^{-\rho n} \left(\rho n \ln \gamma - 2\right) \tag{38}$$

We are now ready to report the two key results from this section.

Theorem A.1. *If $\gamma > 1$, starting from $\rho_0 = 0$ NAM converges without overshoot to the solution of $f(\rho^\star) = 0$, where $\rho^\star$ approximates the node infection probability.*

Proof. Finding a solution for Equation (35) is equivalent to detecting a root of Equation (36). If $\gamma > 1$ and $f(0) \times f''(0) > 0$ it follows from Darboux's theorem [15] that starting from $\rho_0 = 0$ NAM converges without any overshoot to the solution. To check the hypothesis of Darboux's theorem note that

$$f(0) = -1 \text{ and } f(1) = \frac{\mu}{\lambda \gamma} > 0,$$

where $\mu, \lambda > 0$ and $\gamma > 1$. In addition,

$$f'(0) = \frac{\mu}{\lambda \gamma} \text{ and } f''(0) = -2 \frac{\mu n \ln \gamma}{\lambda}.$$

The result follows from the fact that $\ln \gamma > 0$, which implies that $\gamma > 1$ and $f(0) \times f''(0) > 0$. □

Theorem A.2. *The expression of ρ_{i+1} at iteration $i + 1$, as a function of ρ_i produced at iteration i, is given by*

$$\rho_{i+1} = \frac{\lambda - \mu \gamma^{-\rho_i n} (\rho_i^2 n \ln \gamma)}{\lambda - \mu \gamma^{-\rho_i n} (\rho_i n \ln \gamma - 1)}. \tag{39}$$

Proof. According to the Newton Approximation Method (NAM),

$$\rho_{i+1} = \rho_i - \frac{f(\rho_i)}{f'(\rho_i)} \tag{40}$$

$$= \rho_i - \frac{\overbrace{\rho_i + \rho_i \frac{\mu}{\lambda}\gamma^{-\rho_i n} - 1}^{\text{from (36)}}}{\underbrace{1 + \frac{\mu}{\lambda}\gamma^{-\rho_i n}(1 - \rho_i n \ln \gamma)}_{\text{from (37)}}} \tag{41}$$

$$= \frac{\rho_i \frac{\mu}{\lambda}\gamma^{-\rho_i n} - (\rho_i \frac{\mu}{\lambda}\gamma^{-\rho_i n})(\rho_i n \ln \gamma) - \rho_i \frac{\mu}{\lambda}\gamma^{-\rho_i n} + 1}{1 + \frac{\mu}{\lambda}\gamma^{-\rho_i n}(1 - \rho_i n \ln \gamma)}$$

$$= \frac{1 - (\frac{\mu}{\lambda}\gamma^{-\rho_i n})(\rho_i^2 n \ln \gamma)}{1 - \frac{\mu}{\lambda}\gamma^{-\rho_i n}(\rho_i n \ln \gamma - 1)} \tag{42}$$

$$= \frac{\lambda - \mu\gamma^{-\rho_i n}(\rho_i^2 n \ln \gamma)}{\lambda - \mu\gamma^{-\rho_i n}(\rho_i n \ln \gamma - 1)} \tag{43}$$

□

A.3 Closed-form approximation for infection probability

Using the precedent approach it is possible to obtain a closed-form expression for an approximation of the infection probability. Numerically, we experimentally found that using only two iterations of NAM is enough to obtain accurate approximations.

The initial value ρ_0 for NAM is key for the generation of accurate results. We consider two initial values $\rho_0 = 0$ and $\rho_0 = 1$ to obtain two approximations of the infection probability. In Appendix A.4 we present a simple heuristic to determine which is the best initial value.

Let $\rho_i^{(0)}$ be the approximate infection probability after i iterations of NAM, with $\rho_0 = 0$. Then,

$$\begin{aligned} \rho_1^{(0)} &= \frac{\lambda}{\lambda + \mu}, \\ \rho_2^{(0)} &= \frac{\lambda - \mu\gamma^{-\rho_1 n}\left(\rho_1^2 n \ln \gamma\right)}{\lambda - \mu\gamma^{-\rho_1 n}\left(\rho_1 n \ln \gamma - 1\right)} \end{aligned} \tag{44}$$

where $\rho_1 = \rho_1^{(0)}$.

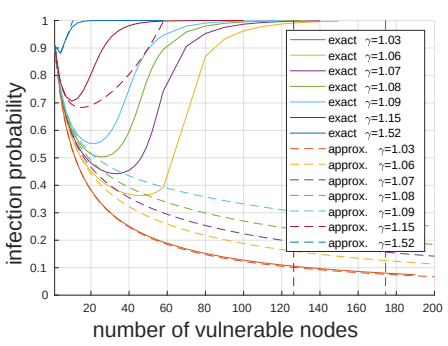
Similarly, let $\rho_i^{(1)}$ be the approximate infection probability after i iterations of NAM, with $\rho_0 = 1$,

$$\rho_1^{(1)} = \frac{\lambda - \mu\gamma^{-n}\left(n \ln \gamma\right)}{\lambda - \mu\gamma^{-n}\left(n \ln \gamma - 1\right)}, \tag{45}$$

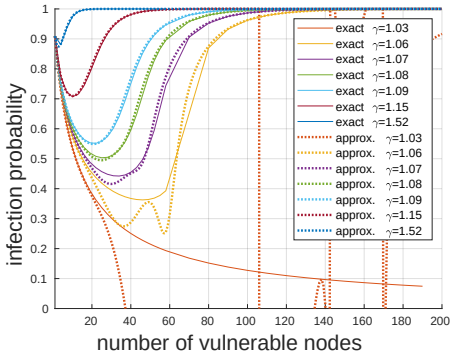
and with $\rho_2^{(1)}$ given by Equation (44), replacing ρ_1 by $\rho_1^{(1)}$.

A.4 Heuristic to set Initial value

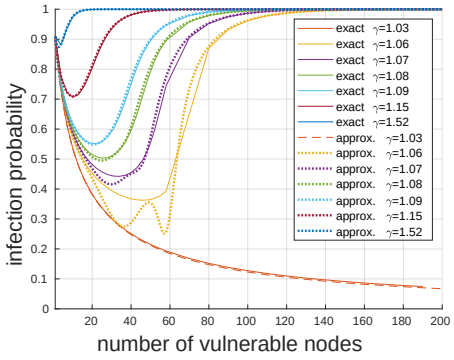
As indicated in Appendix A.3, the accuracy of NAM is very dependent on the considered initial condition. We have shown that to produce tractable closed-form expressions for the infection probability, we can consider two initial conditions that simplify the resulting expressions, namely $\rho_0 = 0$ and $\rho_0 = 1$. In what follows, we indicate an heuristic to



(a) $\rho_0 = 0$



(b) $\rho_0 = 1$



(c) ρ_0 has given by (48)

Fig. 11. Infection probability, obtained through NAM: (a) initial condition $\rho_0 = 0$; (b) $\rho_0 = 1$; and (c) initial condition chosen by the proposed heuristic.

choose between those two initial conditions. The heuristic is inspired by the numerical results presented in Figure 11(a) and Figure 11(b) for $\rho_0 = 0$ and $\rho_0 = 1$, respectively, with the same parameters as in Figure 8 ($\mu = 1$, $\Lambda = 10$ and $\lambda = \Lambda/N$). Note that $\rho_0 = 1$ typically produces good approximations, except when the obtained values evaluate to quantities beyond the range of interest which varies between 0 and 1. In those cases, the approximation through $\rho_0 = 0$ produces accurate estimates.

The discussion in the previous paragraph motivates the following heuristic. First, evaluate the infection probability considering the initial value $\rho_0 = 1$. If the resulting expression to estimate the infection probability produces a value greater than 1 or less than 0, then switch to $\rho_0 = 0$. We denote by $\mathring{\rho}_2(N)$ the infection probability obtained through that simple heuristic,

$$\mathring{\rho}_2(N) = \begin{cases} \rho_2^{(1)}(N), & \text{if } 0 \leq \rho_2^{(1)}(N) \leq 1 \\ \rho_2^{(0)}(N), & \text{otherwise.} \end{cases} \tag{46}$$

Figure 11(c) illustrates the behavior of the proposed heuristic for $\rho \geq 1.06$. In the considered example, $\rho_0 = 1$ produced accurate results except when $\gamma = 1.03$. In the latter case, as shown in Figure 11(b), setting $\rho_0 = 1$ produces results that

are outside the range [0,1]. Hence, for $\gamma = 1.03$ we should set $\rho_0 = 0$ which again produces accurate results as shown in Figure 11(a).

In what follows, we refine the considered heuristic in order to contemplate scenarios such as those corresponding to $\gamma = 1.03$ in the considered setup. To that aim, note that in the numerical examples presented above, when $\gamma \geq 1.06$, large values of N produce an infection probability close to 1, which in turn favor NAM with initial condition $\rho_0 = 1$ as opposed to $\rho_0 = 0$. Accordingly, when $\gamma \leq 1.03$ the infection probability is decreasing with respect to N in the range of interest, favoring $\rho_0 = 0$ irrespectively of N.

Let $\bar{\rho}_2^{(z)}(N)$ be the value of NAM at its second iteration, under initial condition z, if the produced value is in the range between 0 and 1, for all $n \leq N$, and $-\infty$ otherwise. Then,

$$\bar{\rho}_2^{(z)}(N) = \begin{cases} \rho_2^{(z)}(N), & \text{if } 0 \leq \rho_2^{(z)}(N) \leq 1 \\ & \text{and } \bar{\rho}_2^{(z)}(N-1) \neq -\infty, \\ -\infty, & \text{otherwise,} \end{cases} \tag{47}$$

where z, $0 \leq z \leq 1$, is the initial value for ρ_0. Equation (47) explicitly sets the dependence of ρ on $N = n + 1$ (in Lemma 6.1 such dependency was implicit). According to Equation (47), if NAM produces results out of the range [0,1], for a given value of $\tilde{N}$, then $\bar{\rho}_2^{(z)}(N) = -\infty$ for $N \geq \tilde{N}$. The refined heuristic is given by,

$$\bar{\rho}(N) = \max\left(\bar{\rho}_2^{(0)}(N), \bar{\rho}_2^{(1)}(N)\right). \tag{48}$$

Figure 11(c) illustrates the approximation obtained considering the refined heuristic. As shown in Figure 11(c), the refined heuristic captures the fact that $\rho_0 = 0$ should be chosen for $\gamma = 1.03$. We have evaluated the refined heuristic under different configurations (not reported in the paper), and observed that it captured the right initial condition under all the considered examples.

B MULTIPLICATIVE VERSUS ADDITIVE INFECTION MODELS

Next, we further discuss the relationship between additive and multiplicative infection models. First, note that with a logarithmic change of variables, namely, letting $\lambda = \log \tilde{\lambda}$ and $\gamma = \log \tilde{\gamma}$, we have

$$\lambda + \gamma d = \log\left(\tilde{\lambda}\tilde{\gamma}^d\right). \tag{49}$$

The equation above allows us to relate the infection rates under the additive and multiplicative models. A similar idea has been considered in [20], wherein the authors rely on geometric programming for epidemic control after replacing summations by products.

Throughout this paper, we considered the multiplicative model under the assumption that $\gamma > 1$. As argued in Section 3.5, it is always possible to set $\gamma > 1$, as far as time units are conveniently normalized. In the remainder of this appendix, we briefly discuss an interpretation of the model under $\gamma < 1$, which is out of the scope of this paper but may be of interest on its own.

If $\gamma < 1$, the multiplicative model can be interpreted as follows. The infection occurs if the external attacker infects a node *and* all neighbors infect that node as well. The "and" comes in as the multiplication of probabilities, assuming that the respective events are all independent.

Note that under the additive model, a node may be infected externally *or* by any of its neighbors. In that case, the "or" comes in as the addition. In particular, the aggregation of independent Poisson processes is also a Poisson process with rate equal to the sum of the rates of the independent processes.

As in this work we consider the setup wherein the infection rate increases as the number of infected neighbors grows, we assume $\gamma > 1$. In this setting, the multiplicative model is contrasted against the additive model in Section 3.5.

C ALTERNATIVE DERIVATION OF BINOMIAL APPROXIMATION

Zhang and Moura [77] describe an alternative approach to derive a result similar in spirit to Lemma 6.1 taking into account the *most-probable state* $\boldsymbol{x} \in \mathcal{X}$.

Approximation by most-probable state $\boldsymbol{x} \in \mathcal{X}$: Next, we consider an alternative approach to approximate the probability that a node is infected in steady-state. Recall that the steady-state probability of state $\boldsymbol{x}$, $\pi(\boldsymbol{x})$, is given by Equations (6), (7) and (8). In order to approximate the probability that a node is infected in steady-state, Zhang and Moura [77] leverage the notion of *most-probable state* $\boldsymbol{x}^\star = (x_1^\star, x_2^\star, \ldots, x_N^\star)$, with $\boldsymbol{x}^\star = \arg\max_{\boldsymbol{x}\in\mathcal{X}} \pi(\boldsymbol{x})$.

Then, Theorem 4.1 from [77], constitutes an alternative derivation of the binomial approximation. For completeness, we reproduce the theorem below.

Theorem C.1. *If $\tilde{\pi}(\boldsymbol{x}^\star) \gg \tilde{\pi}(\boldsymbol{x})$, $\forall \boldsymbol{x} \in \mathcal{X} \setminus \boldsymbol{x}^\star$, then*

$$P(x_k = 1) \approx \left(1 + \frac{\mu}{\lambda(N)\gamma^{m_k^\star}}\right)^{-1}, \tag{50}$$

where

$$m_k^\star = \sum_{j=1}^{N} a_{k,j} x_j^\star. \tag{51}$$

The proof of Theorem 4.1 in [77] consists of rewriting the steady-state distribution (6)-(7) in light of the Boltzmann distribution. From Equation (7), note that $\tilde{\pi}(\boldsymbol{x}) = e^{H(\boldsymbol{x})}, \quad \boldsymbol{x} \in \mathcal{X}$ where $H(\boldsymbol{x}) = \mathbf{1}^T\boldsymbol{x} \log\left(\frac{\lambda}{\mu}\right) + \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \log\gamma$. Then, Equation (50) is obtained from the relation between $\tilde{\pi}(\boldsymbol{x})$ and the Boltzmann distribution [39].

In Equation (50), $m_k^\star$ is the number of infected neighbors of node k in the most-probable configuration. $N^\star$, determined by Equation (19) in Lemma 6.1, is directly related to $m_k^\star$ determined by Equation (50). Equation (19) is obtained from Equation (50) replacing $m_k^\star$ by $N^\star$.

D EXACT MODEL SOLUTION AND APPROXIMATIONS FOR FULLY CONNECTED NETWORK TOPOLOGY

D.1 General solution

Under the fully connected network topology with N nodes, the number of infected nodes is characterized by a birth-death process. The state of the process is the number of infected nodes. The rate from state i to state $i-1$ equals $i\mu$, as any of the i nodes may recover, for $i = 1, \ldots, N$. The rate from state i to state $i+1$, in turn, depends on whether we consider the multiplicative or the additive model. We denote by $(N-i)\tilde{\Lambda}_i$ the rate from state i to state $i+1$ (see Figure 12). Then, the stationary steady state solution of the system is the classical solution to a birth death Markov chain,

$$\pi_k = \pi_0 \prod_{i=1}^{k} \frac{\tilde{\Lambda}_{i-1}}{i\mu}, \quad k = 1, 2, \ldots, N \tag{52}$$

and

$$\pi_0 = \frac{1}{1 + \sum_{k=1}^{N} \prod_{i=1}^{k} \frac{(N-i)\tilde{\Lambda}_{i-1}}{i\mu}}. \tag{53}$$

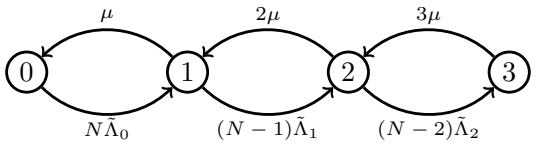


Fig. 12. Markov chain for the fully connected network topology when $N = 3$.

D.2 Multiplicative model

Under the multiplicative model,

$$\tilde{\Lambda}_i = \lambda\gamma^i \tag{54}$$

Therefore,

$$\pi_k = \pi_0 \prod_{i=1}^{k} \frac{(N-i+1)\lambda\gamma^{i-1}}{i\mu}, \quad k = 1, 2, \ldots, N \tag{55}$$

$$= \pi_0 \left(\frac{\lambda}{\mu}\right)^k \frac{\prod_{i=1}^{k}(N-i+1)\gamma^{i-1}}{\prod_{i=1}^{k} i} \tag{56}$$

$$= \pi_0 \left(\frac{\lambda}{\mu}\right)^k \binom{N}{k} \prod_{i=1}^{k} \gamma^{i-1} \tag{57}$$

It follows that

$$\pi_k = \pi_0 \left(\frac{\lambda}{\mu}\right)^k \binom{N}{k} \gamma^{\sum_{i=1}^{k}(i-1)} \tag{58}$$

$$= \pi_0 \left(\frac{\lambda}{\mu}\right)^k \binom{N}{k} \gamma^{k(k-1)/2} \tag{59}$$

In addition,

$$\pi_0 = \frac{1}{1 + \sum_{k=1}^{N} \left(\frac{\lambda}{\mu}\right)^k \binom{N}{k} \gamma^{k(k-1)/2}} \tag{60}$$

and

$$Z = \frac{1}{\pi_0}. \tag{61}$$

The equations above are in agreement with (13).

D.3 Additive model

Under the additive model,

$$\tilde{\Lambda}_i = \lambda + i\gamma \tag{62}$$

Therefore, for $k = 1, 2, \ldots, N$,

$$\pi_k = \pi_0 \prod_{i=1}^{k} \frac{(N-i+1)(\lambda + (i-1\gamma)}{i\mu} \tag{63}$$

$$= \pi_0 \left(\frac{1}{\mu}\right)^k \frac{\prod_{i=1}^{k}(N-i+1)(\lambda+(i-1)\gamma)}{k!} \tag{64}$$

$$= \pi_0 \left(\frac{1}{\mu}\right)^k \binom{N}{k} \prod_{i=1}^{k}(\lambda+(i-1)\gamma), \tag{65}$$

and

$$\pi_0 = \frac{1}{1 + \sum_{k=1}^{N} \left(\frac{1}{\mu}\right)^k \binom{N}{k} \prod_{i=1}^{k}(\lambda+(i-1)\gamma)}. \tag{66}$$

Recall from (11) that

$$\rho(N) = \frac{\mathbb{E}(I)}{N} = \tag{67}$$

$$= \frac{1}{N} \sum_{k=0}^{N} k\pi_0 \left(\frac{1}{\mu}\right)^k \binom{N}{k} \prod_{i=1}^{k}(\lambda+(i-1)\gamma), \tag{68}$$

where

$$\mathbb{E}(I) = \sum_{\iota=0}^{N} \iota\pi_\iota \tag{69}$$

is the expected number of infected nodes.

D.4 NIMFA approximation under additive model

The direct and exact solution of the infection model (65) involves a product not expressed in closed form. In part, this occurs because the model solution requires the characterization of the infection probability of each node conditioned on the states of neighboring nodes. The states of the neighboring nodes, in turn, is captured through the expected value of a product of random variables. In this section, we rely on a mean-field (MF) approximation referred to as N-interwinded MF approximation (NIMFA) to compute the fraction of infected nodes. The approximation consists of replacing the expectation of the product of random variables by the product of their expectations [30, 50, 60].

Let $X_{j,0}(t)$ and $X_{j,I}(t)$ be two indicator random variables equal to 1 if node j is healthy or infected, respectively. Accordingly, let $\pi_{j,0}(t)$ and $\pi_{j,I}(t)$ be the probabilities that node j is healthy or infected at time t, respectively. Note that

$$\mathbb{E}(X_{j,0}(t)) = \pi_{j,0}(t), \tag{70}$$

$$\mathbb{E}(X_{j,I}(t)) = \pi_{j,I}(t). \tag{71}$$

The time change of $\mathbb{E}(X_{j,I}(t))$ is given by

$$\frac{d\pi_{j,I}(t)}{dt} = -\mu\pi_{j,I}(t) + \mathbb{E}\left(X_{j,0}(t)\left(\lambda + \gamma \sum_{k\neq j} X_{k,I}(t)\right)\right)$$

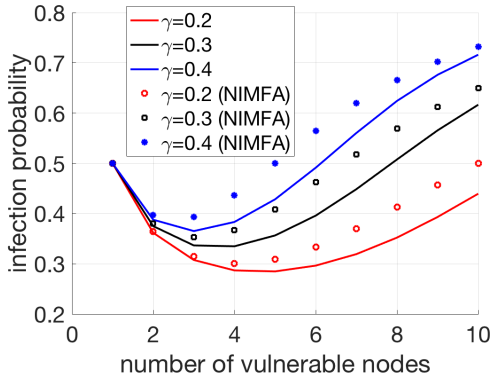


Fig. 13. Additive model under complete graph topology: comparing exact solution against NIMFA approximation. We let $\lambda = 1/N$ and $\mu = 1$, varying γ between 0.2, 0.3 and 0.4.

As mentioned above, the NIMFA approximation consists of replacing the expectation of the product of random variables by the product of their expectations. Let $r_j(t)$ be the endogenous infection rate towards node j by its neighbors, at time t. Under NIMFA, $r_j(t)$ is approximated as follows,

$$r_j(t) = \gamma \sum_{k \neq j} X_{k,I}(t) \approx \sum_{k \neq j} \gamma \pi_{k,I}(t)$$

which yields

$$\frac{d\pi_{j,I}(t)}{dt} = -\mu \pi_{j,I}(t) + \left(\lambda + \sum_{k \neq j} \gamma \pi_{k,I}(t) \right) \pi_{j,0}(t). \tag{72}$$

Leveraging the symmetry between nodes, we let

$$\lim_{t \to \infty} \pi_{j,I}(t) = \rho.$$

Then, in stationary regime it follows from (72) that

$$\begin{aligned} 0 &= -\mu\rho + \big(\lambda + (N-1)\gamma\rho\big)(1-\rho) \\ &= -(N-1)\gamma\rho^2 - \big((\lambda+\mu) - (N-1)\gamma\big)\rho + \lambda. \end{aligned}$$

Whenever the equation above admits a root between 0 and 1, it is given by

$$\rho(N) = \frac{(\lambda+\mu) - (N-1)\gamma \pm \sqrt{\Delta}}{2(1-N)\gamma} \tag{73}$$

where

$$\Delta = \big((\lambda+\mu) - (N-1)\gamma\big)^2 + 4(N-1)\gamma\lambda.$$

In particular, if $\lambda = 0$ and $(N-1)\gamma > \mu$ the solution is given by

$$\rho(N; \lambda = 0) = 1 - \frac{1}{\gamma(N-1)/\mu}.$$

which is in agreement with [30].

Figure 13 shows the infection probability ρ as a function of N, letting $\mu = 1$ and $\lambda = 1/N$. The full lines are obtained through the exact solution of the Markov chain (Equation (68)) whereas the circles are obtained through the NIMFA approximation (Equation (73)) for $\gamma = 0.2$, 0.3 and 0.4, respectively. As in the case of the multiplicative model considered in the remainder of this paper, the infection probability first decreases and then increases as the number of vulnerable nodes grows. In addition, the NIMFA approximation captures well the behavior of the exact MC solution, allowing to find the number of vulnerable nodes that minimizes the infection probability. It is also worth noting that the NIMFA model overestimates the infection probability, which is in agreement with [13, 61] although the assumptions of those related works do not account for exogenous infections (see also [12, 55]). A more careful analysis of the conditions under which NIMFA overestimates the infection probability, as well as of the connections between the exact MC solution and NIMFA, are left as subject for future work, noting that a detailed discussion about NIMFA accuracy under general topologies can be found at [50].

D.5 NIMFA approximation under multiplicative model

Next, we consider the multiplicative model under the NIMFA approximation. Using the same terminology as in the previous section, the time change of $\mathbb{E}(X_{j,I}(t))$ is given by

$$\frac{d\pi_{j,I}(t)}{dt} = -\mu\pi_{j,I}(t) + \mathbb{E}\left(X_{j,0}(t)\lambda\gamma^{\sum_{k\neq j} X_{k,I}(t)}\right).$$

As mentioned above, the NIMFA approximation consists of replacing the expectation of the product of random variables by the product of their expectations. In the scenario of the multiplicative model, we also consider an additional approximation, which consists of replacing the expectation $\mathbb{E}\left(\gamma^{X_{k,I}(t)}\right)$ by $\gamma^{\mathbb{E}(X_{k,I}(t))}$.

In summary, we consider the following two approximations:

- **(A1) independence approximation:** replace the expectation of the product of random variables by the product of expectations;
- **(A2) functional approximation:** replace the expectation of a function, $\mathbb{E}\left(\gamma^{X}\right)$, by a function of the expectation, $\gamma^{\mathbb{E}(X)}$.

Under the two approximations above, the endogenous infection rate towards node j by its neighbors, at time t, is given by

$$r_j(t) = \gamma^{\sum_{k\neq j} X_{k,I}(t)} \approx \gamma^{\sum_{k\neq j} \pi_{k,I}(t)}$$

which yields

$$\frac{d\pi_{j,I}(t)}{dt} = -\mu\pi_{j,I}(t) + \left(\lambda\gamma^{\sum_{k\neq j} \pi_{k,I}(t)}\right)\pi_{j,0}(t). \tag{74}$$

Leveraging the symmetry between nodes, we let

$$\lim_{t\to\infty} \pi_{j,I}(t) = \rho. \tag{75}$$

Then, in stationary regime it follows from (74) that

$$0 = -\mu\rho + \left(\lambda\gamma^{(N-1)\rho}\right)(1-\rho)$$

Therefore,

$$\rho = \frac{1}{1 + \mu\left(\lambda\gamma^{(N-1)\rho}\right)^{-1}}. \tag{76}$$

The above derivation indicates that the NIMFA approximation provides an alternative derivation and rationale to approximation (35), referred to as the *binomial approximation* in this paper.

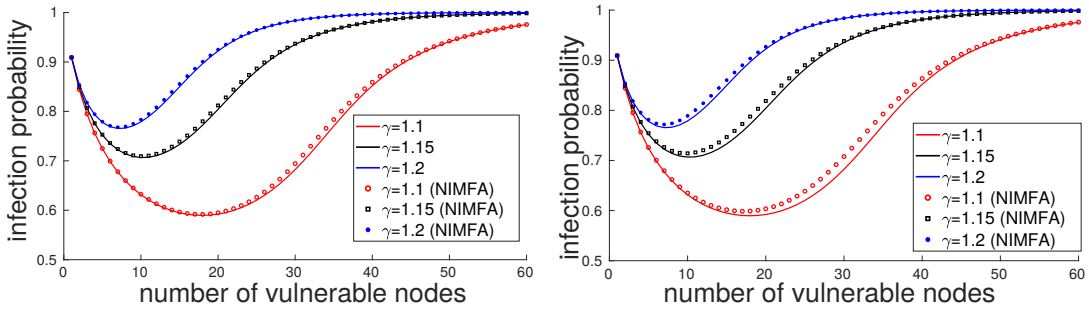
Alternatively, if we consider only approximation (A1), but not (A2), the infection probability is given by the root of the following equation,

$$0 = -\mu\rho + \lambda(1-\rho)\Big(\gamma\rho + (1-\rho)\Big)^{(N-1)}. \tag{77}$$

In the numerical experiments that follow, we indicate that (76) typically provides better approximations than (77). This occurs as we empirically observed that (76) slightly overestimates the infection probability. This, in turn, is in agreement with [13, 61]. Then, approximation (A2) serves as a correction. Indeed, it follows from Jensen inequality that

$$\mathbb{E}\left(\gamma^X\right) \geq \gamma^{\mathbb{E}(X)}, \qquad \gamma > 1, \quad 0 \leq X \leq 1. \tag{78}$$

The inequality above implies that approximation (A2) favors a reduction in the infection probability, and that together (A1) and (A2) balance out to produce better approximations through (76) when compared against (77).



(a) approximation (A1) and (A2) (b) approximation (A1)

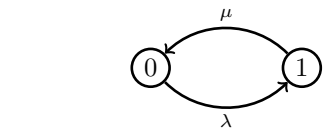
Fig. 14. Multiplicative model under complete graph topology: comparing exact solution against NIMFA approximation, letting $\mu = 1$ and $\Lambda = 10$: (a) under approximations (A1) and (A2) and (b) under approximation (A1).

Figure 14 shows the infection probability ρ as a function of N, letting $\mu = 1$, $\Lambda = 10$ and $\lambda = \Lambda/N$. The full lines are obtained through the exact solution of the Markov chain (Equations (59)-(61)) whereas the stars, squares and circles are obtained through the NIMFA approximation for $\gamma = 1.1$, 1.15 and 1.2, respectively. Equation (76) is used to obtain Figure 14(a), under approximations (A1) and (A2), and Equation (77) is used to obtain Figure 14(b), under approximation (A1).

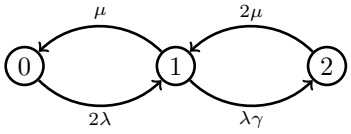
As in Appendix D.4, the NIMFA approximation captures the behavior of the exact MC solution, allowing to find the number of vulnerable nodes that minimizes the infection probability. In addition, the NIMFA approximation again overestimates the infection probability. Applying approximation (A2) on top of (A1) favors a correction of the overshooting, as evidenced by the closer agreement between the exact MC solution and the approximations in Figure 14(a) when compared against Figure 14(b).

E EXACT MODEL SOLUTION AND APPROXIMATIONS FOR BIPARTITE NETWORK TOPOLOGIES

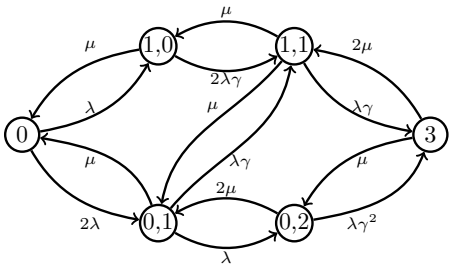
Next, we consider the solution of the model for bipartite network topologies. Figure 15 shows the Markov chains corresponding to the proposed epidemic model accounting for up to 4 vulnerable nodes, assuming a bipartite network



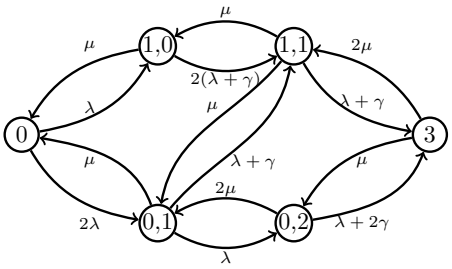
(a) One vulnerable node



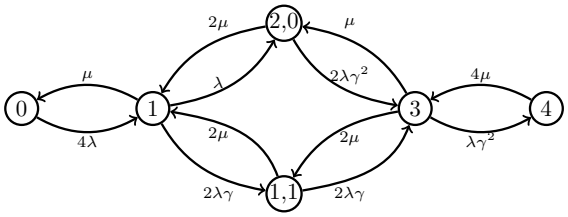
(b) Two vulnerable nodes



(c) Three vulnerable nodes multiplicative model



(d) Three vulnerable nodes: additive model



(e) Four vulnerable nodes: multiplicative model

Fig. 15. Markov chains for the bipartite topology with up to 4 vulnerable nodes

topology. The Markov chains leverage symmetry in the bipartite graph. In what follows, we solve the corresponding Markov chains.

To appreciate the analysis that follows through a very simple example, we start considering the case of a single vulnerable node. The solution in this simple case, as well as in the case of two vulnerable nodes, is in agreement with the full topology considered so far. Then, we consider three and four nodes, indicating the specifics of the role played by topology on the model solution.

E.1 One vulnerable node

In this case, states 0 and 1 correspond to the vulnerable node being susceptible and infected, respectively (see Figure 15(a)). The corresponding steady state probabilities are π_0 and π_1. Then,

$$\pi_0\lambda = \mu\pi_1$$

$$\pi_0 + \pi_1 = 1$$

Therefore,

$$\pi_0 = \frac{\mu}{\lambda+\mu}, \quad \pi_1 = \frac{\lambda}{\lambda+\mu}$$

Alternatively, we can rely on results from the detailed state space introduced in Section 5.3 to derive the same results. In this appendix, we refer to a state of the Markov chain considered in the rest of this paper as a *detailed state*, and to

the lumped states considered in this appendix simply as *states*. To each state i we associated its corresponding class of symmetric detailed states. The symmetric detailed states in each class have all the same steady state probability. Let ν_i be the number of symmetric detailed states in the class of state i. The steady state probability of each of those symmetric detailed states equals π_i/ν_i. It follows from (6)-(8) that

$$\tilde{\pi}_i = \nu_i \left(\frac{\lambda}{\mu}\right)^{1x^T} \gamma^{x^T A x/2} \tag{79}$$

$$\pi_i = \frac{\tilde{\pi}_i}{\sum\limits_{\forall j} \tilde{\pi}_j} \tag{80}$$

Then,

$$\tilde{\pi}_0 = \left(\frac{\lambda}{\mu}\right)^0 \gamma^0 = 1$$

$$\tilde{\pi}_1 = \frac{\lambda}{\mu}^1 \gamma^0 = \frac{\lambda}{\mu}$$

and

$$\pi_0 = \frac{\tilde{\pi}_0}{\sum\limits_{\forall j} \tilde{\pi}_j} = \frac{1}{1+\frac{\lambda}{\mu}} = \frac{\mu}{\lambda+\mu}$$

$$\pi_1 = \frac{\tilde{\pi}_1}{\sum\limits_{\forall j} \tilde{\pi}_j} = \frac{\frac{\lambda}{\mu}}{1+\frac{\lambda}{\mu}} = \frac{\lambda}{\lambda+\mu}$$

The infection probability is given by $\rho = \pi_1$.

E.2 Two vulnerable nodes

E.2.1 Multiplicative model. The Markov chain corresponding to the multiplicative model with two nodes is shown in Figure 15(b). In that case, state i corresponds to i infected nodes in the network, $i = 0, 1, 2$. The corresponding steady state probabilities are π_i. Then,

$$\pi_0 2\lambda = \mu\pi_1$$

$$\pi_1 \lambda\gamma = 2\mu\pi_2$$

$$\pi_0 + \pi_1 + \pi_2 = 1$$

Therefore,

$$\pi_0 + \frac{\pi_0 2\lambda}{\mu} + \left(\frac{\pi_0 2\lambda}{\mu}\right)\frac{\lambda\gamma}{2\mu} = 1$$

and

$$\pi_0 = \frac{1}{1 + \frac{2\lambda}{\mu} + \frac{\lambda^2\gamma}{\mu^2}} = \frac{\mu^2}{\mu^2 + 2\lambda\mu + \lambda^2\gamma}$$
$$\pi_1 = \frac{2\lambda\mu}{\mu^2 + 2\lambda\mu + \lambda^2\gamma}$$
$$\pi_2 = \frac{\lambda^2\gamma}{\mu^2 + 2\lambda\mu + \lambda^2\gamma}$$

As discussed in the previous section, those results can also be similarly obtained using the detailed state space introduced in Section 5.3. Considering the same terminology as the one introduced in the previous section (see Equations (79)-(80)),

$$\tilde{\pi}_0 = 1\frac{\lambda}{\mu}^0\gamma^0 = 1$$
$$\tilde{\pi}_1 = 2\frac{\lambda}{\mu}^1\gamma^0 = \frac{2\lambda}{\mu}$$
$$\tilde{\pi}_2 = 1\frac{\lambda}{\mu}^2\gamma^1 = \frac{\lambda^2\gamma}{\mu^2}$$

Then,

$$\pi_0 = \frac{\tilde{\pi}_0}{\sum\limits_{\forall j}\tilde{\pi}_j} = \frac{1}{1 + \frac{2\lambda}{\mu} + \frac{\lambda^2\gamma}{\mu^2}} = \frac{\mu^2}{\mu^2 + 2\lambda\mu + \lambda^2\gamma}$$
$$\pi_1 = \frac{\tilde{\pi}_1}{\sum\limits_{\forall j}\tilde{\pi}_j} = \frac{2\frac{\lambda}{\mu}}{1 + \frac{2\lambda}{\mu} + \frac{\lambda^2\gamma}{\mu^2}} = \frac{2\lambda\mu}{\mu^2 + 2\lambda\mu + \lambda^2\gamma}$$
$$\pi_2 = \frac{\tilde{\pi}_2}{\sum\limits_{\forall j}\tilde{\pi}_j} = \frac{\frac{\lambda^2\gamma}{\mu^2}}{1 + \frac{2\lambda}{\mu} + \frac{\lambda^2\gamma}{\mu^2}} = \frac{\lambda^2\gamma}{\mu^2 + 2\lambda\mu + \lambda^2\gamma}$$

The infection probability is given by $\rho = (\pi_1 + 2\pi_2)/2$,

$$\rho = \frac{\mathbb{E}(I)}{2} = \frac{\frac{\lambda}{\mu} + \frac{\lambda^2\gamma}{\mu^2}}{1 + \frac{2\lambda}{\mu} + \frac{\lambda^2\gamma}{\mu^2}}. \tag{81}$$

E.2.2 Additive model. The Markov chain corresponding to the additive model with two nodes is obtained from the one shown in Figure 15(b), replacing the rate from state 1 to state 2 from $\lambda\gamma$ by $\lambda + \gamma$. In that case, state i corresponds to i infected nodes in the network, $i = 0, 1, 2$. The corresponding steady state probabilities are π_i. Then,

$$\pi_0 2\lambda = \mu\pi_1$$
$$\pi_1(\lambda + \gamma) = 2\mu\pi_2$$
$$\pi_0 + \pi_1 + \pi_2 = 1$$

Therefore,

$$\pi_0 = \frac{1}{1 + \frac{2\lambda}{\mu} + \frac{2\lambda}{\mu}\frac{\gamma+\lambda}{2\mu}}$$
$$\pi_1 = \frac{\frac{2\lambda}{\mu}}{1 + \frac{2\lambda}{\mu} + \frac{\lambda}{\mu}\frac{\gamma+\lambda}{\mu}}$$
$$\pi_2 = \frac{\frac{\lambda}{\mu}\frac{\gamma+\lambda}{\mu}}{1 + \frac{2\lambda}{\mu} + \frac{\lambda}{\mu}\frac{\gamma+\lambda}{\mu}}$$

The infection probability is given by $\rho = (\pi_1 + 2\pi_2)/2$,

$$\rho = \frac{\mathbb{E}(I)}{2} = \frac{\frac{\lambda}{\mu} + \frac{\lambda(\gamma+\lambda)}{\mu^2}}{1 + \frac{2\lambda}{\mu} + \frac{\lambda(\lambda+\gamma)}{\mu^2}}. \tag{82}$$

E.3 Three vulnerable nodes

In this case, the bipartite topology is composed of two subgraphs. One subgraph contains one node, and the other contains two nodes. The fact that the number of nodes in each subgraph is distinct breaks symmetry, and requires us to keep track of two state variables at four states, where the first state variable characterizes the state of the subgraph comprised of a single node, and the second state variable corresponds to the state of the other subgraph. State 0 corresponds to 0 infected nodes. State (1,0) corresponds to one isolated node being infected in the subgraph comprised of a single node. State (0,1), in contrast, corresponds to one node being infected in the subgraph comprised of two nodes. As pointed out above, we need to distinguish states (1,0) and (0,1) due to symmetry breaking. Similarly, states (1,1) and (0,2) correspond to one node in each subgraph being infected, and two nodes in the same subgraph being infected, respectively. Finally, state 3 corresponds to all nodes being infected.

E.3.1 Multiplicative model. The Markov chain corresponding to the multiplicative model is shown in Figure 15(c). The flow balance equations are given as follows,

$$\pi_0 3\lambda = \mu(\pi_{0,1} + \pi_{1,0}) \tag{83}$$
$$(\pi_{1,0} 2\gamma + \pi_{0,1}(1+\gamma))\lambda = 2\mu(\pi_{0,2} + \pi_{1,1}) \tag{84}$$
$$\pi_{0,2}\lambda\gamma^2 + \pi_{1,1}\lambda\gamma = 3\mu\pi_3 \tag{85}$$
$$\pi_{1,0}(\mu + 2\lambda\gamma) = \pi_0\lambda + \pi_{1,1}\mu \tag{86}$$
$$\pi_{0,1}(\mu + \lambda + \lambda\gamma) = \pi_0 2\lambda + \pi_{1,1}\mu + \pi_{0,2}2\mu \tag{87}$$

and

$$\pi_0 + \pi_{1,0} + \pi_{0,1} + \pi_{0,2} + \pi_{1,1} + \pi_3 = 1 \tag{88}$$

We also let

$$\pi_1 = \pi_{1,0} + \pi_{0,1} \tag{89}$$
$$\pi_2 = \pi_{0,2} + \pi_{1,1} \tag{90}$$

The first three equations above are obtained by considering the balance of flow in and flow out between the four layers of states, i.e., accounting for sets of states 0, $\{0, (1,0), (0,1)\}$, and $\{0, (1,0), (0,1), (1,1), (0,2)\}$, respectively. The following two equations correspond to flow in balancing flow out in states (1,0) and (0,1), respectively.

Using the detailed state space introduced in Section 5.3, we compute the steady state probabilities. Considering the same terminology as the one introduced in the previous section (see (79)-(80)),

$$\begin{aligned}
\tilde{\pi}_0 &= 1\frac{\lambda}{\mu}^0 \gamma^0 = 1\\
\tilde{\pi}_{1,0} &= 1\frac{\lambda}{\mu}^1 \gamma^0 = \frac{\lambda}{\mu}\\
\tilde{\pi}_{0,1} &= 2\frac{\lambda}{\mu}^1 \gamma^0 = 2\frac{\lambda}{\mu}\\
\tilde{\pi}_{1,1} &= 2\left(\frac{\lambda}{\mu}\right)^2 \gamma^1 = 2\frac{\lambda^2\gamma}{\mu^2}\\
\tilde{\pi}_{0,2} &= 1\left(\frac{\lambda}{\mu}\right)^2 \gamma^0 = \frac{\lambda^2}{\mu^2}\\
\tilde{\pi}_3 &= 1\left(\frac{\lambda}{\mu}\right)^3 \gamma^2 = \frac{\lambda^3\gamma^2}{\mu^3}
\end{aligned}$$

Then,

$$\pi_i = \frac{\tilde{\pi}_i}{1 + \frac{3\lambda}{\mu} + \frac{\lambda^2(1+2\gamma)}{\mu^2} + \frac{\lambda^3\gamma^2}{\mu^3}}, \qquad i = 0, 1, 2, 3, 4. \tag{91}$$

It can be readily verified that the solution above satisfies (83)-(90).

The infection probability of a uniformly selected node is given by

$$\begin{aligned}
\rho &= \frac{1}{3}\left(\pi_{10} + +\pi_{11} + \pi_3\right) + \frac{2}{3}\left(\frac{\pi_{01} + \pi_{11}}{2} + \pi_{02} + \pi_3\right)\\
&= \frac{\mathbb{E}(I)}{3} = \frac{1}{3}\left(\frac{3\frac{\lambda}{\mu} + 4\frac{\lambda^2\gamma}{\mu^2} + 2\frac{\lambda^2}{\mu^2} + 3\frac{\lambda^3\gamma^2}{\mu^3}}{1 + \frac{3\lambda}{\mu} + \frac{\lambda^2(1+2\gamma)}{\mu^2} + \frac{\lambda^3\gamma^2}{\mu^3}}\right).
\end{aligned}$$

E.3.2 Additive Model. The Markov chain corresponding to the additive model is shown in Figure 15(d). The flow balance equations are given as follows,

$$\begin{aligned}
\pi_0 3\lambda &= \mu(\pi_{0,1} + \pi_{1,0})\\
2\pi_{1,0}(\lambda + \gamma) + \pi_{0,1}(2\lambda + \gamma) &= 2\mu(\pi_{0,2} + \pi_{1,1})\\
\pi_{0,2}(\lambda + 2\gamma) + \pi_{1,1}(\lambda + \gamma) &= 3\pi_3\mu\\
\pi_{1,0}(\mu + 2\lambda + 2\gamma) &= \pi_0\lambda + \pi_{1,1}\mu\\
\pi_{0,1}(\mu + 2\lambda + \gamma) &= \pi_0 2\lambda + \mu(\pi_{1,1} + 2\pi_{0,2})
\end{aligned}$$

and

$$\pi_0 + \pi_{1,0} + \pi_{0,1} + \pi_{0,2} + \pi_{1,1} + \pi_3 = 1$$

Symbolically solving the system of equations above is a daunting task, which evidences the benefits of the multiplicative model for which there are closed form expressions for the stationary state probabilities. Nonetheless, with the

help of the Matlab symbolic solver, we are able to express all quantities in closed form. Letting $\mu = 1$,

$$\pi_i = \frac{\tilde{\pi}_i}{Z}$$
$$Z = \sum_{i=0}^{5} \zeta_i \lambda^i$$

where

$$\begin{aligned}
\tilde{\pi}_0 &= 14\lambda^2 + (21\gamma + 19)\lambda + \zeta_0 \\
\tilde{\pi}_{10} &= 30\lambda^3 + (39\gamma + 43)\lambda^2 + (33\gamma + 12\gamma^2 + 9)\lambda \\
\tilde{\pi}_{01} &= 12\lambda^3 + (24\gamma + 14)\lambda^2 + (12\gamma^2 + 12\gamma + 18)\lambda \\
\tilde{\pi}_{02} &= 60\lambda^4 + (138\gamma + 102)\lambda^3 + \\
&\quad + (170\gamma + 102\gamma^2 + 42)\lambda^2 + (36\gamma + 70\gamma^2 + 24\gamma^3)\lambda \\
\tilde{\pi}_{11} &= 12\lambda^4 + (30\gamma + 6)\lambda^3 + (10\gamma + 24\gamma^2 + 6)\lambda^2 + \\
&\quad + (4\gamma^2 + 6\gamma^3)\lambda \\
\tilde{\pi}_3 &= \zeta_5\lambda^5 + (84\gamma + 36)\lambda^4 + (98\gamma + 108\gamma^2 + 16)\lambda^3 + \\
&\quad + (30\gamma + 88\gamma^2 + 60\gamma^3)\lambda^2 + (12\gamma^2 + 26\gamma^3 + 12\gamma^4)\lambda
\end{aligned}$$

and

$$\begin{aligned}
\zeta_5 &= 24 \\
\zeta_4 &= 84\gamma + 108 \\
\zeta_3 &= 108\gamma^2 + 266\gamma + 166 \\
\zeta_2 &= 60\gamma^3 + 214\gamma^2 + 273\gamma + 119 \\
\zeta_1 &= 12\gamma^4 + 56\gamma^3 + 110\gamma^2 + 102\gamma + 46 \\
\zeta_0 &= 8\gamma^2 + 15\gamma + 9
\end{aligned}$$

Contrasting the equation above against the solution to the multiplicative model (see Equation (91)), we note that the multiplicative model is instrumental to analyze and study general topologies.

E.4 Four vulnerable nodes

In this case, we have a bipartite network comprised of two subgraphs with two nodes each. Except when there are two infected nodes in the network, it suffices to keep track of the number of infected nodes in the network. Therefore, state i corresponds to i infected nodes, for $i = 0, 1, 3, 4$. When $i = 2$, we need to distinguish between two states: (2,0) and (1,1). At state (2,0), we have two nodes infected in the same subgraph, noting that the identity of the subgraph is irrelevant. At state (1,1), in contrast, we have two nodes infected, each node in a distinct subgraph.

E.4.1 Multiplicative model. The Markov chain corresponding to the multiplicative model is shown in Figure 15(e). The flow balance equations are given as follows,

$$\pi_0 4\lambda = \mu\pi_1 \tag{92}$$

$$\pi_1(2\lambda\gamma + \lambda) = 2\mu(\pi_{2,0} + \pi_{1,1}) \tag{93}$$

$$\pi_{2,0}2\lambda\gamma^2 + \pi_{1,1}2\lambda\gamma = 3\mu\pi_3 \tag{94}$$

$$\pi_3\lambda\gamma^2 = 4\mu\pi_4 \tag{95}$$

$$\pi_1 2\lambda\gamma + \pi_3 2\mu = (2\mu + 2\lambda\gamma)\pi_{1,1} \tag{96}$$

and

$$\pi_0 + \pi_1 + \pi_{2,0} + \pi_{1,1} + \pi_3 + \pi_4 = 1 \tag{97}$$

We also let

$$\pi_2 = \pi_{2,0} + \pi_{1,1} \tag{98}$$

The first four equations above are obtained by considering the balance of flow in and flow out of the sets of states 0, $\{0, 1\}$, $\{0, 1, (2, 0), (1, 1)\}$ and $\{0, 1, (2, 0), (1, 1), 3\}$, respectively. Equation (96) corresponds to balancing flow in and flow out from state (1,1).

Using the detailed state space introduced in Section 5.3, we compute the steady state probabilities. Considering the same terminology as the one introduced in the previous section (see (79)-(80)),

$$\tilde{\pi}_0 = 1\left(\frac{\lambda}{\mu}\right)^0 \gamma^0 = 1 \tag{99}$$

$$\tilde{\pi}_1 = 4\left(\frac{\lambda}{\mu}\right)^1 \gamma^0 = 4\frac{\lambda}{\mu} \tag{100}$$

$$\tilde{\pi}_{1,1} = 4\left(\frac{\lambda}{\mu}\right)^2 \gamma^1 = 4\frac{\lambda^2\gamma}{\mu^2} \tag{101}$$

$$\tilde{\pi}_{2,0} = 2\left(\frac{\lambda}{\mu}\right)^2 \gamma^0 = 2\frac{\lambda^2}{\mu^2} \tag{102}$$

$$\tilde{\pi}_2 = \frac{2\lambda^2(1 + 2\gamma)}{\mu^2} \tag{103}$$

$$\tilde{\pi}_3 = 4\left(\frac{\lambda}{\mu}\right)^3 \gamma^2 = 4\frac{\lambda^3\gamma^2}{\mu^3} \tag{104}$$

$$\tilde{\pi}_4 = 1\left(\frac{\lambda}{\mu}\right)^4 \gamma^4 = \frac{\lambda^4\gamma^4}{\mu^4} \tag{105}$$

Then,

$$\pi_i = \frac{\tilde{\pi}_i}{1 + \frac{4\lambda}{\mu} + \frac{2\lambda^2(1+2\gamma)}{\mu^2} + \frac{4\lambda^3\gamma^2}{\mu^3} + \frac{\lambda^4\gamma^4}{\mu^4}}, \tag{106}$$

for $i \in \{0, 1, (1, 1), (0, 2), 2, 3, 4\}$, where $\tilde{\pi}_i$ is given by (99)-(105).

The infection probability is given by

$$\rho = \frac{1}{4}\left(\frac{4\frac{\lambda}{\mu} + 8\frac{\lambda^2\gamma}{\mu^2} + 4\frac{\lambda^2}{\mu^2} + 12\frac{\lambda^3\gamma^2}{\mu^3} + 4\frac{\lambda^4\gamma^4}{\mu^4}}{1 + \frac{4\lambda}{\mu} + \frac{2\lambda^2(1+2\gamma)}{\mu^2} + \frac{4\lambda^3\gamma^2}{\mu^3} + \frac{\lambda^4\gamma^4}{\mu^4}}\right).$$

E.4.2 Additive model. Next, we present the solution to the four node bipartite topology under the additive model. The solution serves to further evidence the simplicity of the multiplicative model. As in the previous section, with the help of the Matlab symbolic solver, we are able to express all quantities in closed form. Letting $\mu = 1$,

$$\pi_i = \frac{\tilde{\pi}_i}{Z}$$
$$Z = \sum_{i=0}^{5} \zeta_i \lambda^i$$

where

$$\begin{aligned}
\tilde{\pi}_0 &= 3\lambda + \zeta_0 \\
\tilde{\pi}_1 &= 4\lambda(3\lambda + \zeta_0) \\
\tilde{\pi}_{20} &= 2\lambda(2\gamma^2 + 5\gamma\lambda + 3\lambda^2 + 3\lambda) \\
\tilde{\pi}_{11} &= 4\lambda(4\gamma^2 + 8\gamma\lambda + 3\gamma + 3\lambda^2 + 3\lambda) \\
\tilde{\pi}_3 &= 4\lambda(4\gamma^3 + 12\gamma^2\lambda + 2\gamma^2 + 11\gamma\lambda^2 + 6\gamma\lambda + 3\lambda^3 + 3\lambda^2) \\
\tilde{\pi}_4 &= (2\gamma + \lambda)\tilde{\pi}_3/4
\end{aligned}$$

and

$$\begin{aligned}
\zeta_5 &= 3 \\
\zeta_4 &= 17\gamma + 15 \\
\zeta_3 &= 56\gamma + 34\gamma^2 + 30 \\
\zeta_2 &= 66\gamma + 62\gamma^2 + 28\gamma^3 + 30 \\
\zeta_1 &= 32\gamma + 28\gamma^2 + 20\gamma^3 + 8\gamma^4 + 15 \\
\zeta_0 &= 5\gamma + 3
\end{aligned}$$

Comparing the equations above against those derived in Section E.3.2, we note that due to symmetry the solution of the four node topology is much simpler than that for three nodes. Nonetheless, further contrasting the equations above against the multiplicative model (see Equation (106)), we again appreciate that the multiplicative model is instrumental to analyze and study general topologies.

E.5 General number of nodes

Next, we consider a bipartite graph with $N = 2(\tilde{N} - 1)$ nodes, with $\tilde{N} - 1$ nodes in each partition. A naive solution to compute the infection probability involves a Markov chain with state space cardinality of $\tilde{N}^2$, as each partition can have from 0 up to $\tilde{N} - 1$ infected nodes. If $N = 4$, this amounts to 9 states. Nonetheless, further leveraging the problem symmetry we can lump the state space, e.g., leading to the 6 state Markov chain in the case of four nodes (Figures 15(c) and 15(d) show the lumped MCs for the multiplicative and additive models, respectively).

E.5.1 Lumped state space cardinality. Next, we compute the cardinality of the lumped state space. The state space is divided into layers, where each layer ℓ corresponds to a given number of infected nodes, $\ell = 0, 1, \ldots, 2(\tilde{N}-1)$. At layer ℓ, the number of states equals the number of ways to throw ℓ balls into 2 indistinguishable bins (corresponding to the two partitions of the bipartite graph), where each bin can contain up to $\tilde{N}-1$ balls. Let $\mathcal{B}(\ell)$ be the number of states at layer ℓ. Then, $\mathcal{B}(\ell)$ is given by the Gaussian binomial coefficient,

$$\mathcal{B}(\ell) = \left[q^{\ell}\right] \binom{\tilde{N}+1}{2}_q$$

where $\left[q^{\ell}\right] P$ denotes the coefficient of q^{ℓ} in polynomial P and

$$\binom{\tilde{N}+1}{2}_q = \frac{(1-q^{\tilde{N}+1})(1-q^{\tilde{N}})}{(1-q)(1-q^2)}.$$

Let $|\Omega_B|$ be the cardinality of the state space of the lumped model corresponding to the bipartite topology. Then,

$$\begin{aligned} |\Omega_B| &= \sum_{\ell=0}^{2(\tilde{N}-1)} \mathcal{B}(\ell) = \binom{\tilde{N}+1}{2}_q \Bigg|_{q=1} \\ &= (\tilde{N}+1)\tilde{N}/2. \end{aligned} \tag{107}$$

The rationale goes as follows. If there are $\tilde{N}-1$ nodes per partition, there are $\tilde{N}(\tilde{N}-1)/2$ ways to configure the number of infections in the top and bottom partitions, restricted by the number of infected nodes in the top partition being strictly larger than the number of infected nodes in the bottom one. In addition, there are $\tilde{N}$ configurations in which the number of infected nodes in both partitions is the same. Together, $\tilde{N}(\tilde{N}-1)/2 + \tilde{N}$ equals (107). If $N = 20,000$, for instance, then $|\Omega_B| = 50,015,001$ and the original state space has cardinality 100,020,001.

E.5.2 Additive model. Next, we consider the additive model. To simplify presentation, we account for the unlumped version of the model, wherein there are $\tilde{N}^2$ states. Each state (i, j) corresponds to i infected nodes in the top partition and j infected nodes in the bottom one. Then, the positive entries of the infinitesimal generator matrix $\boldsymbol{Q}$ are given by $q_{(i,j),(k,l)}$,

$$q_{(i,j),(k,l)} = \begin{cases} ((\tilde{N}-1)-i)\lambda + \\ \quad +j((\tilde{N}-1)-i)\gamma, & \text{if } k = i+1 \text{ and } j = l \\ ((\tilde{N}-1)-j)\lambda + \\ \quad +i((\tilde{N}-1)-j)\gamma, & \text{if } i = k \text{ and } l = j+1 \\ i\mu, & \text{if } i = k+1 \text{ and } j = l \\ j\mu, & \text{if } j = l+1 \text{ and } i = k \\ 0, & \text{otherwise} \end{cases}$$

where i, j, k and l are all greater than or equal to zero, and strictly smaller than $\tilde{N}$. Then, the steady state solution is given by the standard flow balance equations,

$$\pi \boldsymbol{Q} = 0, \quad \sum_i \sum_j \pi_{ij} = 1, \tag{108}$$

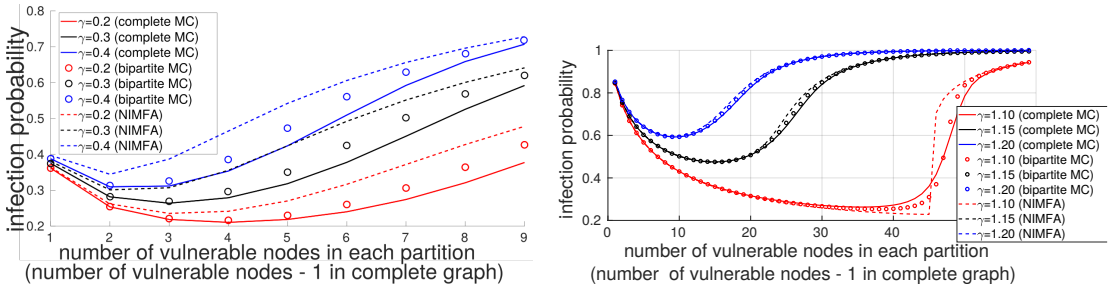


Fig. 16. NIMFA approximation: NIMFA is insensitive to the network topology being a bipartite graph or a complete graph. The exact solution of the Markov chain model indicates that there is a slight difference between the exact solution of the the two under (a) additive model and (b) multiplicative model. We let $\Lambda = 10$, $\mu = 1$ and γ vary between 1.1, 1.15 and 1.2 under the multiplicative model and $\Lambda = 1$, $\mu = 1$ and γ vary between 0.2, 0.3 and 0.4 under the additive model.

where π is the vector of stationary state probabilities. In particular, if $\tilde{N} = 2$ and $\tilde{N} = 3$, the solution above is in agreement with Appendices E.2.2 and E.4.2, respectively.

E.5.3 Multiplicative model. Next, we consider the multiplicative model. As in the previous section, to simplify presentation, we account for the unlumped version of the model, wherein there are $\tilde{N}^2$ states. Each state (i, j) corresponds to i infected nodes in the top partition and j infected nodes in the bottom one. Then, it follows from (6)-(8) that

$$\pi_{ij} = \pi_{00} \binom{\tilde{N}-1}{i} \binom{\tilde{N}-1}{j} \left(\frac{\lambda}{\mu}\right)^{i+j} \gamma^{ij}$$

$$\pi_{00} = $$

$$= \frac{1}{\sum_{i=0}^{\tilde{N}-1} \sum_{j=0}^{\tilde{N}-1} \binom{\tilde{N}-1}{i} \binom{\tilde{N}-1}{j} \left(\frac{\lambda}{\mu}\right)^{i+j} \gamma^{ij}}$$

and

$$\mathbb{E}(I) = \sum_{i=0}^{\tilde{N}-1} \sum_{j=0}^{\tilde{N}-1} (i+j)\pi_{ij}$$

$$\rho = \frac{\mathbb{E}(I)}{2(\tilde{N}-1)}.$$

In particular, if $\tilde{N} = 2$ and $\tilde{N} = 3$, the solution above is in agreement with Appendices E.2.1 and E.4.1, respectively.

E.6 NIMFA approximation

For the additive model, the analysis that leads to Equation (73) still holds.

$$\rho(2(\tilde{N}-1)) = \frac{-(\lambda+\mu) + (\tilde{N}-1)\gamma + \sqrt{\Delta}}{2(\tilde{N}-1)\gamma} \tag{109}$$

where

$$\Delta = ((\lambda+\mu) - (\tilde{N}-1)\gamma)^2 + 4(\tilde{N}-1)\gamma\lambda. \tag{110}$$

If $\lambda = 0$ and $(\tilde{N}-1)\gamma > \mu$ the solution is given by

$$\rho(2(\tilde{N}-1);\lambda=0) = 1 - \frac{1}{\gamma(\tilde{N}-1)/\mu}. \tag{111}$$

which is in agreement with [30].

Similarly, under the multiplicative model the NIMFA approximation for bipartite networks is given by

$$\rho(2(\tilde{N}-1)) = \frac{1}{1+\mu\left(\lambda\gamma^{(\tilde{N}-1)}\rho\right)^{-1}}. \tag{112}$$

The equation above is in agreement with (25), noting that all nodes have the same degree equal $\tilde{N}-1$, i.e., $\tilde{d} = \tilde{N}-1$.

Comparing the equations above against those presented in Appendix D, we note that the same equations hold in the two considered scenarios. The fact that the NIMFA equations for the complete and bipartite graphs are the same reflects the fact that NIMFA, in this setup, is insensitive to the specifics of the topology, which are captured only through node degrees. This is due to the fact that NIMFA captures the direct impact of the neighbors of a node, but not second order effects, e.g., due to neighbors of neighbors.

Figure 16 shows the NIMFA solution for the bipartite graph (dotted lines), and contrasts it against the exact solution of complete graphs (full line) and bipartite graphs (circles, squares and dots). Figures 16(a) and 16(b) correspond to the additive and multiplicative models, respectively. In all scenarios, we assume

$$\lambda = \frac{\Lambda}{2(\tilde{N}-1)} \tag{113}$$

where $\tilde{N}-1$ is the number of vulnerable nodes in each partition of the bipartite graphs. In the complete graphs, $\tilde{N}$ is the number of vulnerable nodes in the network. Note that to allow for a comparison between the bipartite and complete topologies, in this scenario we exceptionally assume that λ decays according to (113) under the complete topology rather than $\lambda = \Lambda/\tilde{N}$.

Figure 16 shows that even though NIMFA does not distinguish between the two topologies, in reality there is a gap between the exact solution of the two. Figure 16(a) shows that under the additive model, the infection probability estimated by NIMFA is typically larger than the infection probabilities of bipartite graphs and complete graphs. In addition, the former is typically larger than the latter. In all cases, the infection probability first decreases and then increases as the number of vulnerable nodes grows.

Figure 16(b) shows that under the multiplicative model the infection probability of a bipartite graph with $\tilde{N}-1$ nodes per partition is typically larger than that of a complete graph with $\tilde{N}$ nodes in the intermediary regime wherein the system transitions from being dominated by exogenous infections to being dominated by endogenous infections. The NIMFA approximation overestimates the infection probability of complete and bipartite graphs in that regime, e.g., when $\tilde{N}-1$ varies between 45 and 50 in the setup where $\gamma = 1.1$ the dotted line (NIMFA) is above the full line (complete MC) and the dots (bipartite MC).

F CONDITIONS UNDER WHICH INFECTION PROBABILITY INITIALLY DECREASES AS POPULATION GROWS

One of the key results in the paper relates is the observation that the infection probability may increase as the population of vulnerable nodes grows. Next, we rely on results from the previous sections to establish conditions under which the infection probability decreases when the number of vulnerable nodes grows from one to two. We consider both the

additive and the multiplicative models of infection propagation. Recall that

$$\lambda = \frac{\Lambda}{N}.$$

In all scenarios, we have

$$\rho(1) = \frac{\Lambda}{\Lambda + \mu}$$

Next, we evaluate $\rho(2)$ and establish conditions under which $\rho(2) < \rho(1)$.

F.1 Multiplicative model

Under the multiplicative model, we have from (81), replacing λ by $\Lambda/2$,

$$\rho(2) = \frac{\frac{\Lambda}{2\mu} + \frac{\Lambda^2\gamma}{4\mu^2}}{1 + \frac{\Lambda}{\mu} + \frac{\Lambda^2\gamma}{4\mu^2}} \tag{114}$$

Therefore,

$$\begin{aligned}\rho(1) > \rho(2) &\Rightarrow \frac{\Lambda}{\Lambda+\mu} > \frac{\frac{\Lambda}{2\mu} + \frac{\Lambda^2\gamma}{4\mu^2}}{1 + \frac{\Lambda}{\mu} + \frac{\Lambda^2\gamma}{4\mu^2}} \\ &\Rightarrow \frac{1}{1+\frac{\mu}{\Lambda}}\left(1 + \frac{\Lambda}{\mu} + \frac{\Lambda^2\gamma}{4\mu^2}\right) > \frac{\Lambda}{2\mu} + \frac{\Lambda^2\gamma}{4\mu^2}\end{aligned}$$

In particular, letting $\Lambda = \mu = 1$,

$$\begin{aligned}\rho(1) > \rho(2) &\Rightarrow \frac{1}{2}\left(2 + \frac{\gamma}{4}\right) > \frac{1}{2} + \frac{\gamma}{4} \\ &\Rightarrow \gamma < 4.\end{aligned}$$

Recall that under the multiplicative model we assume $\gamma > 1$. Hence, there is an initial decrease in ρ as N increases from 1 to 2 if $\lambda = \mu = 1$ and $1 < \gamma < 4$.

The illustrative example above serves to indicate conditions under which the infection probability decreases as the number of vulnerable nodes grows. There are a number of other scenarios under which the considered behavior holds, and an extensive analysis of necessary and sufficient conditions is left as subject for future work.

F.2 Additive model

Under the additive model, we have from (82),

$$\rho(2) = \frac{\frac{\Lambda}{2\mu} + \frac{\Lambda(\gamma+\Lambda/2)}{2\mu^2}}{1 + \frac{\Lambda}{\mu} + \frac{\Lambda(\Lambda/2+\gamma)}{2\mu^2}}$$

Therefore,

$$\begin{aligned}\rho(1) > \rho(2) &\Rightarrow \frac{\Lambda}{\Lambda+\mu} > \frac{\frac{\Lambda}{2\mu} + \frac{\Lambda(\gamma+\Lambda/2)}{2\mu^2}}{1 + \frac{\Lambda}{\mu} + \frac{\Lambda(\Lambda/2+\gamma)}{2\mu^2}} \\ &\Rightarrow \frac{1}{1+\frac{\mu}{\Lambda}}\left(1 + \frac{\Lambda}{\mu} + \frac{\Lambda^2/2 + \Lambda\gamma}{2\mu^2}\right) > \frac{\Lambda}{2\mu} + \frac{\Lambda^2/2 + \Lambda\gamma}{2\mu^2}\end{aligned}$$

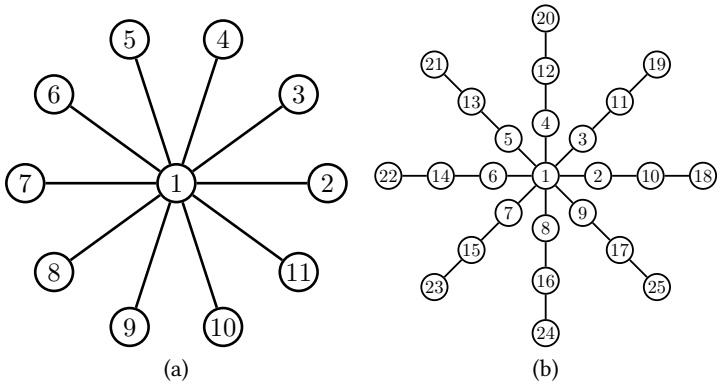


Fig. 17. Two topologies with central hubs: (a) star graph with 11 nodes and (b) star-ring graph with 8 branches and 25 nodes.

In particular, letting $\Lambda = \mu = 1$,

$$\rho(1) > \rho(2) \Rightarrow 1.5 > \gamma$$

Hence, there is an initial decrease in ρ as N increases from 1 to 2 if $\Lambda = \mu = 1$ and $0 < \gamma < 1.5$. As under the multiplicative model, there are a number of other scenarios under which the considered behavior holds, and an extensive analysis of necessary and sufficient conditions is left as subject for future work.

G SIMULATIONS UNDER DIFFERENT TOPOLOGIES

Next, our goal is to investigate the role of network topologies on the spread of epidemics, beyond the complete and bipartite graphs studied so far. To that aim, we use the simulator presented in Section 8. We considered the following four topologies: star, star-ring, tree-cluster and star-cliques (see Figures 17-19). The topologies are further described in the sections that follows, and the parameters used in our simulations are those reported in Table 2. We run simulations for 10,000 time units, which is long enough to estimate the network's steady state. Each configuration was executed three times; in Figures 20-23 we plot the infection probability average with a 95% confidence interval as a function of the number of vulnerable hosts.

G.1 Star and star-ring topologies

In the star topology, all nodes are connected with the central node, as shown in Figure 17(a). In the star-ring topology, each branch is connected with the central node, as shown in Figure 17(b). Those type of topologies are widely used in computer networks, where nodes may be physically interconnected through a central hub or switch, or logically connected to a single central point that controls all communications.

The simulation results of an epidemic process accounting for a strategic attacker on top of a star topology and of a star-ring topology are shown in Figures 20 and 21. We observe that under those two topologies, exogenous infections typically play a more significant role than the endogenous ones. For that reason, the infection probability usually decreases as the number of nodes grows. As nodes are connected only through the central hub, there is not much opportunity for endemic transmissions.

Endogenous infections may play a role depending on the system parameters. For instance, under the star topology, if the endogenous infection rate is high or the uptime is large (last row of Figure 20) we observe a slight increase in the

fraction of endogenously infected nodes as the number of vulnerable nodes increases. In all other considered scenarios, endogenous infections play a negligible role.

G.2 Tree-cluster topology

The $\mathcal{N}$-ary tree-cluster topology is characterized by cliques organized in a tree topology, where each tree node has $\mathcal{N}$ children, as shown in Figure 18. In Figure 18, filled ellipses represent fully connected subgraphs. This model represents a topology where intranets form an hierarchy, e.g., Internet autonomous systems and departments within a company.

Results from our simulations with a strategic attacker on an $\mathcal{N}$-ary tree-cluster topology are reported in Figure 22. The results are similar to those for star topologies, as the node degrees are limited and exogenous infections dominate the epidemic behavior.

G.3 Star-cliques topology

In the star-cliques topology we have a clique of $\mathcal{N}$ core nodes and $\mathcal{N}$ cliques of $\tilde{\mathcal{N}}$ nodes, wherein one of the latter nodes per clique is connected to a single distinct core node, as shown in Figure 19. This topology approximates intranets logically and physically connected through a wide-area network such as the Internet.

The results of our simulations with a strategic attacker on a star-like topology are reported in Figure 23. The results are similar to the scenario of fully-connected topologies (Figure 10): an initial prevalence of exogenous infections is followed by a regime wherein endogenous infections are more relevant.

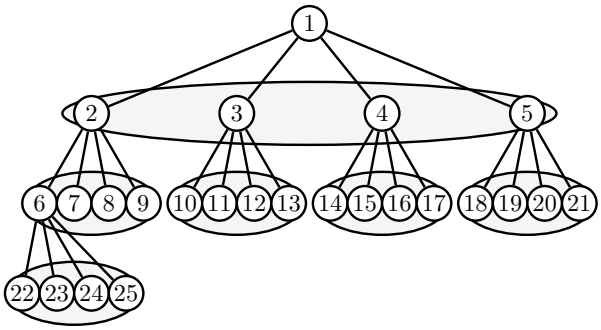


Fig. 18. Tree-cluster network topology: there are $\mathcal{N} = 4$ children per node and a total of $N = 25$ nodes. Filled ellipses represent fully connected subgraphs.

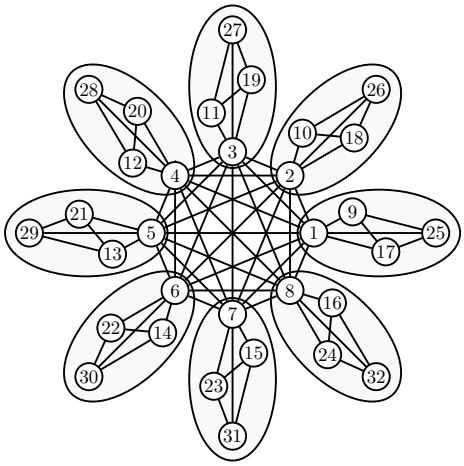


Fig. 19. Star-cliques graph with 8 clusters and 32 nodes.

G.4 Summary

Across all topologies, we observed that exogenous infections have an important impact on the epidemic behavior. The prevalence of endogenous infections as the number of nodes increases is dependent on nodes having sufficient out-degree to allow the infection to spread. In topologies where node degrees are small relative to topology size (e.g., star-ring topologies), endogenous infections have a negligible effect, indicating the relevance of capturing the characteristics of exogenous infections as indicated throughout this work.

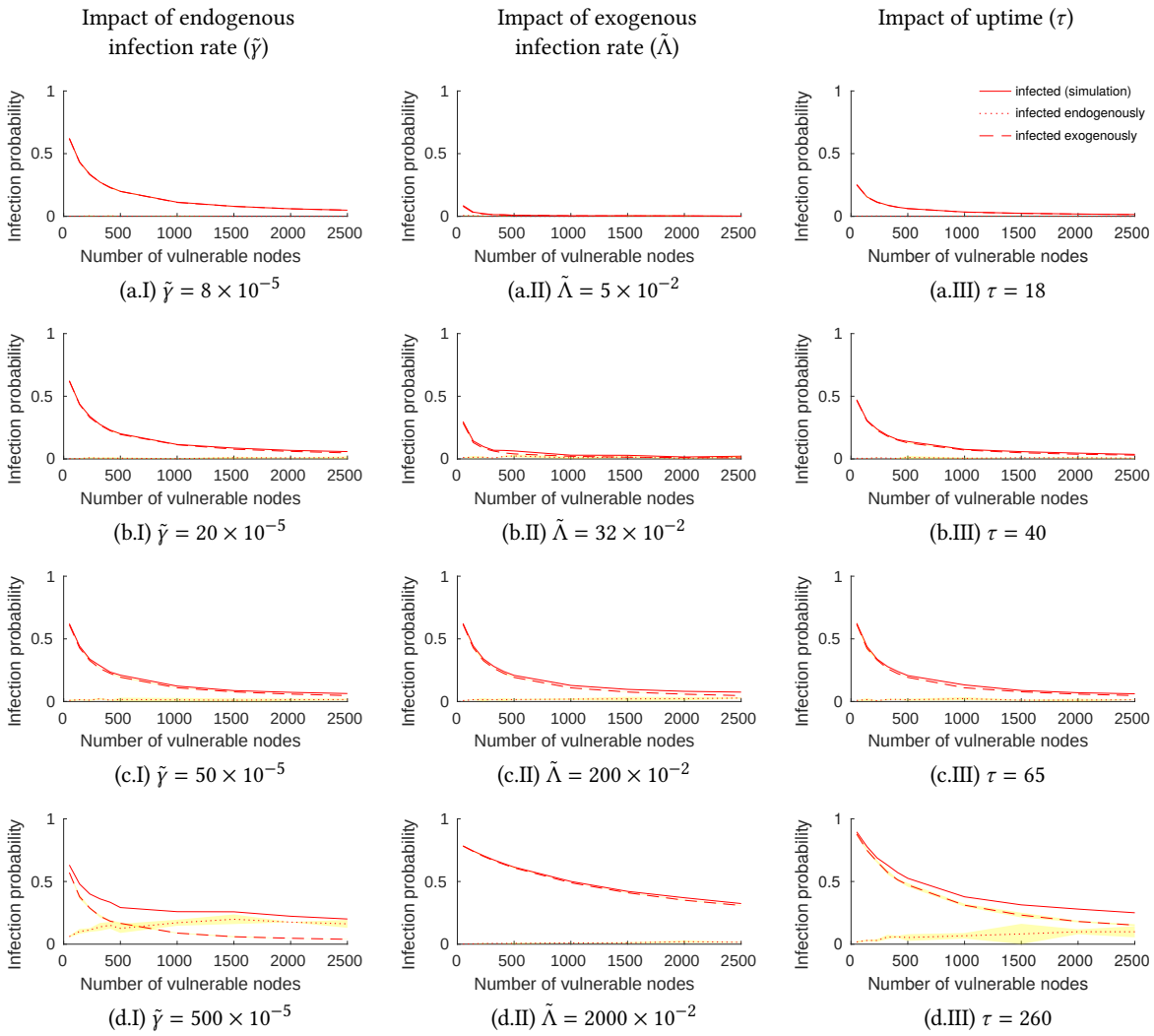


Fig. 20. Outcome of the simulation experiments in a star network (Figure 17(a)), under the action of the *Mirai Botnet* in presence of a strategic attacker. The reference values of the simulator parameters are: $\Lambda = 1500$, $\tilde{\gamma} = 5 \times 10^{-5}$, $\tilde{\Lambda} = 2 \times 10^{-2}$ and $\tau = 65$. Model parameters are shown in Table 3.

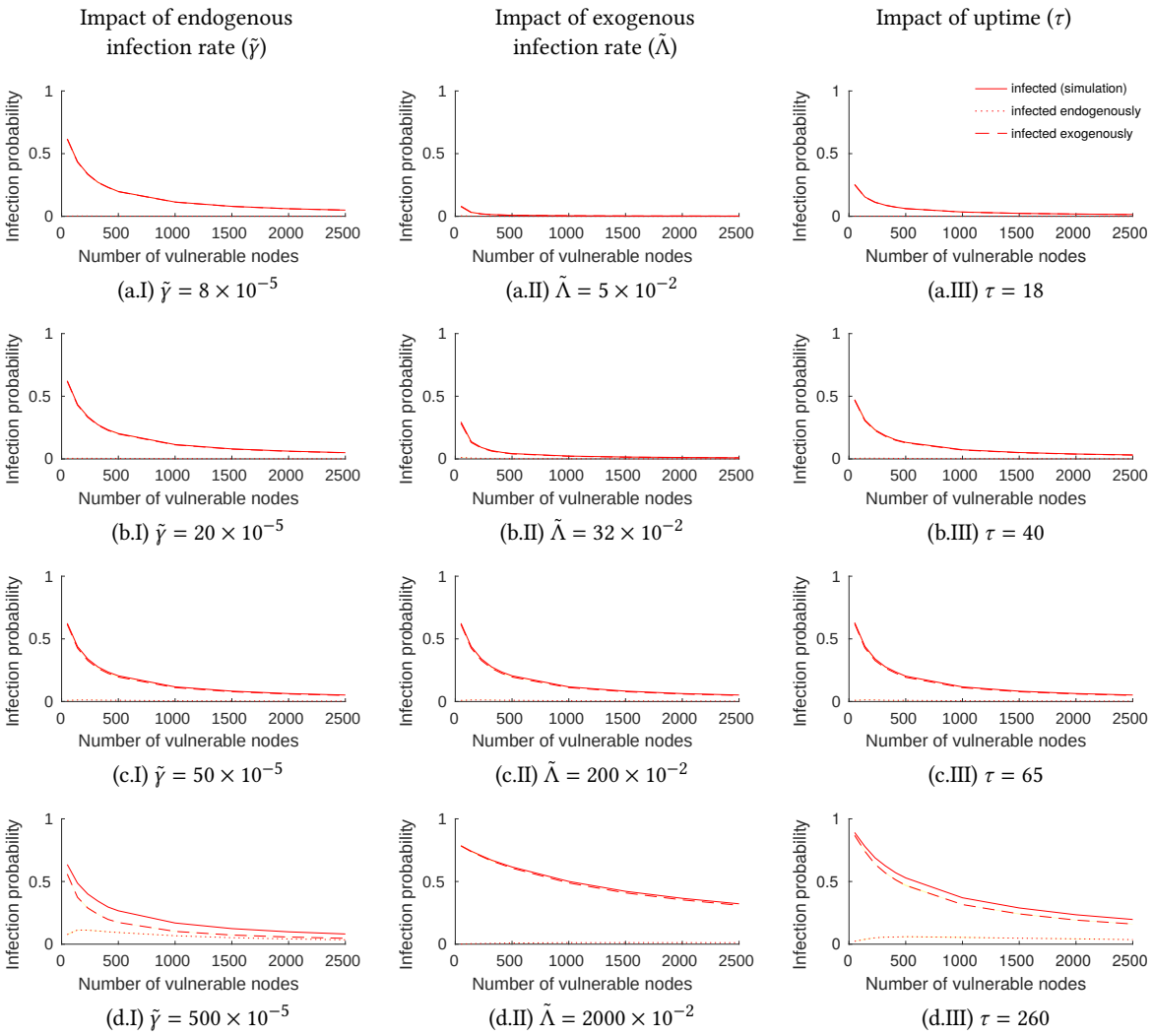


Fig. 21. Outcome of the simulation experiments in a star-ring network with 7 branches (Figure 17(b)), under the action of the *Mirai Botnet* in presence of a strategic attacker. The reference values of the simulator parameters are: $\tilde{\Lambda} = 2 \times 10^{-2}$ and $\tau = 65$.

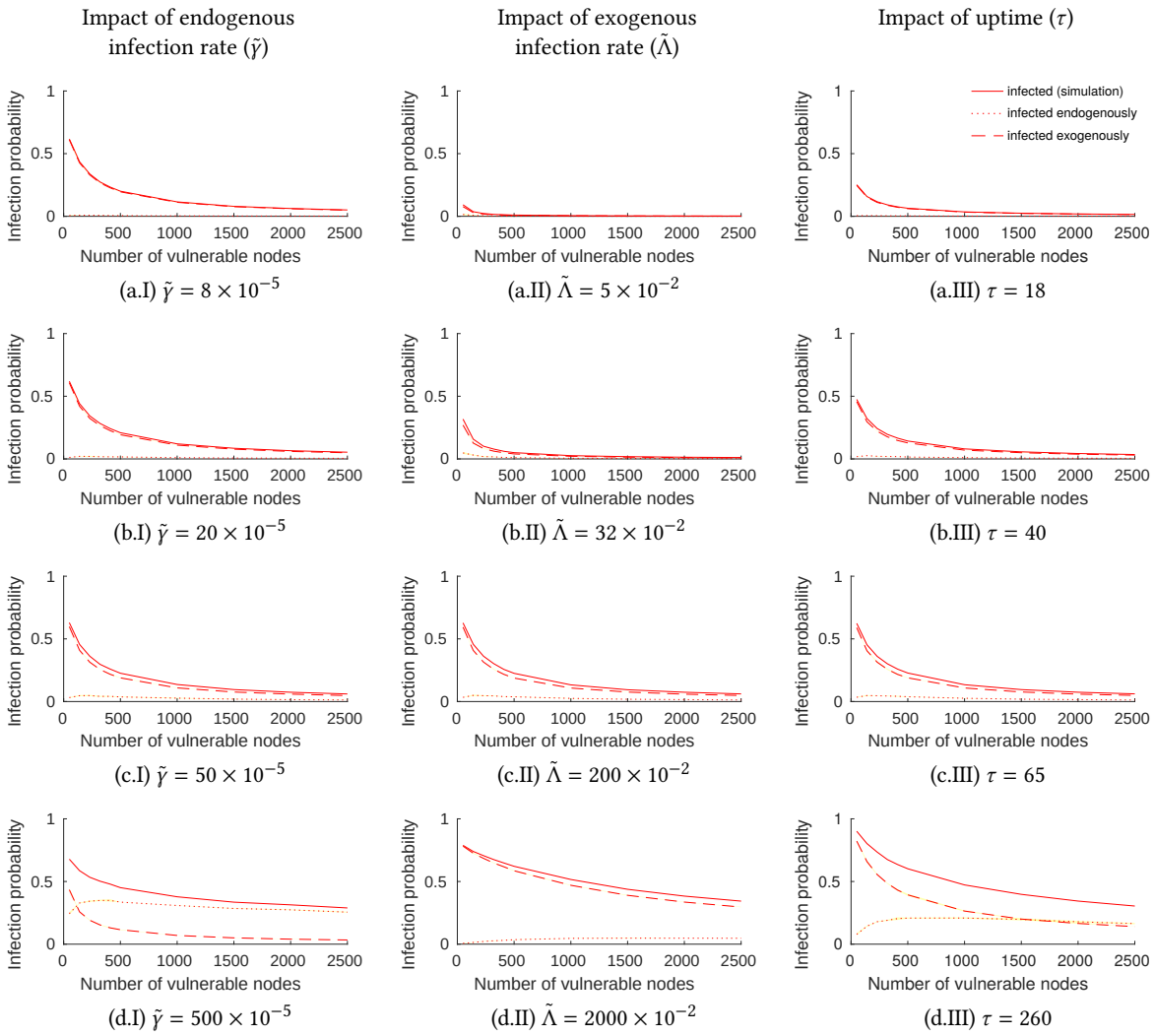


Fig. 22. Outcome of the simulation experiments in a tree-cluster network (Figure 18), under the action of the *Mirai Botnet* in presence of a strategic attacker. The reference values of the simulator parameters are: $\tilde{\Lambda} = 2 \times 10^{-2}$ and $\tau = 65$.

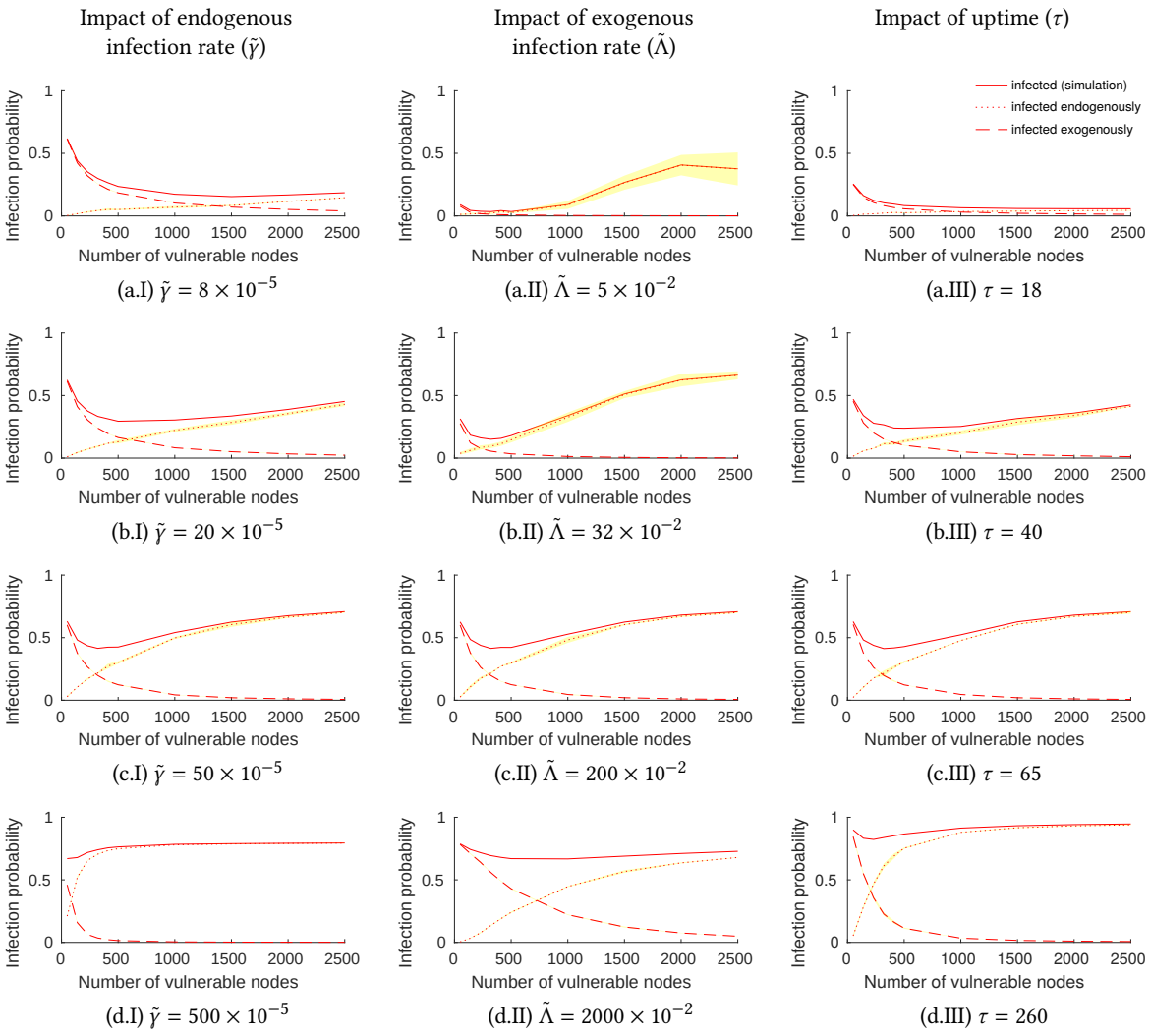


Fig. 23. Outcome of the simulation experiments in a star-cliques network (Figure 19), under the action of the *Mirai Botnet* in presence of a strategic attacker. The reference values of the simulator parameters are: $\tilde{\Lambda} = 2 \times 10^{-2}$ and $\tau = 65$.