MODELING AND PREDICTING NETWORK EPIDEMICS BASED ON RANDOM WALKS WITH LOCAL PROTECTION EFFECTS

Ronald Chiesse de Souza

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Sistemas e Computação.

Orientador: Daniel Ratton Figueiredo

Rio de Janeiro
Dezembro de 2023

MODELING AND PREDICTING NETWORK EPIDEMICS BASED ON RANDOM WALKS WITH LOCAL PROTECTION EFFECTS

Ronald Chiesse de Souza

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Orientador: Daniel Ratton Figueiredo

Aprovada por: Prof. Daniel Ratton Figueiredo
Prof. Valmir Carneiro Barbosa
Prof. Fabrício Murai Ferreira
Prof. Pedro Henrique González Silva
Prof. Don Fred Towsley

RIO DE JANEIRO, RJ – BRASIL
DEZEMBRO DE 2023

Chiesse de Souza, Ronald

Modeling and Predicting Network Epidemics based on Random Walks with Local Protection Effects/Ronald Chiesse de Souza. – Rio de Janeiro: UFRJ/COPPE, 2023.

XIII, 68 p.: il.; 29, 7cm.

Orientador: Daniel Ratton Figueiredo

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2023.

Referências Bibliográficas: p. 60 – 67.

1. network epidemics. 2. protection effects. 3. disease-behavior dynamics. 4. agent-based model. 5. fluid model. 6. markov chains. 7. random walks. I. Ratton Figueiredo, Daniel. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

À minha família.

Acknowledgements

I thank God first and foremost.

I am also profoundly grateful to my family, for being decisively supportive throughout this journey.

Many thanks to my advisor, Daniel, for supporting this work since its very beginning with rich and enlightening discussions.

Last, I enormously appreciate the time and attention devoted by the evaluation committee—to whom I am sincerely thankful—to review this work.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

MODELING AND PREDICTING NETWORK EPIDEMICS BASED ON RANDOM WALKS WITH LOCAL PROTECTION EFFECTS

Ronald Chiesse de Souza

Dezembro/2023

Orientador: Daniel Ratton Figueiredo

Programa: Engenharia de Sistemas e Computação

Efeitos de proteção (PrE) são contramedidas que indivíduos tomam (e.g. uso de máscara) contra epidemias. Tal é a mudança no padrão de mobilidade, influenciada ativamente pela noção de risco e passivamente pelas restrições geoespaciais. Esta tese investiga aspectos teóricos da estrita interação entre mobilidade ciente de risco e estrutura de rede em processos epidêmicos, a fim de *prever* seu resultado. Propomos um modelo simples de agentes móveis para epidemias SIS de tempo contínuo em redes de graus não correlacionados. PrE surgem pelo enviesamento dos agentes para locais seguros. Nossas principais contribuições: (i) Um preditor preciso baseado em EDO que incorpora explicitamente informações de estrutura e proteção em um sistema ordens de grandeza menor que o tamanho da rede; (ii) A primeira evidência teórica de um impacto estrutural chave: maior *heterogeneidade* estimula a propagação; (iii) Com destaque, encontramos um regime especial onde epidemias em redes com *distribuição de grau arbitrária* são previstas em $\Theta(1)$ não por um sistema, mas por uma *única* equação; (iv) Limiares de proteção para estados estacionários livres de doença, sob taxas de caminhada e transmissibilidade tanto constantes quanto arbitrariamente altas; (v) Uma lei pela qual um passeio aleatório simples (SRW) pode ser combinado com a função de grau para produzir *qualquer momento desejado da distribuição de grau de uma rede*. Entre outras consequências, respondemos assim à questão sobre o valor esperado que um SRW se aproxima ao superestimar o grau médio de uma rede. Embora centrados em teoria, nossos resultados indicam que a disponibilidade sem precedentes de formas de rastrear locais de risco torna a mobilidade inteligente uma alternativa plausível e de menor impacto a quarentenas. Acreditamos que nossos resultados também podem contribuir para a teoria de interação de partículas.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

MODELING AND PREDICTING NETWORK EPIDEMICS BASED ON RANDOM WALKS WITH LOCAL PROTECTION EFFECTS

Ronald Chiesse de Souza

December/2023

Advisor: Daniel Ratton Figueiredo

Department: Systems Engineering and Computer Science

Protection effects (PrE) denote countermeasures individuals take (e.g. maskwearing) against epidemics. Such is the collective change on mobility patterns, influenced actively by risk-awareness and passively by geospatial constraints. This work investigates theoretical aspects on the strict interplay between risk-aware mobility and network structure in epidemic processes, in order to *predict* their outcome. We propose and formalize a simple agent-based model for continuous-time SIS epidemics on degree-uncorrelated networks. Protective behavior comes through biasing walkers towards safe sites. Our main contributions: (i) we provide an accurate ODE-based predictor that explicitly embodies structural and protective information into a system orders of magnitude smaller than the network size; (ii) the first theoretical evidence of a key structural impact on agent-based network epidemics: larger *heterogeneity* boosts the spreading; (iii) remarkably, we find a special regime under which epidemics on networks with *arbitrary degree distribution* can be accurately predicted in $\Theta(1)$ not by a system but by a *single* ODE equation; (iv) protection thresholds for disease-free steady states under both constant and arbitrarily large rates for walk and transmissibility; (v) interestingly, we find a law through which a simple random walk (SRW) can be conveniently combined with the degree function in order to yield *any desired moment of a network's degree distribution*. Among other consequences, we thereby answer the question on the expected value an SRW approaches in reality when overestimating a network's average degree. While theory-centric, our results indicate that today's unprecedented availability of ways to keep track of risky locations make the development of smart mobility schemes a plausible, lower-impact alternative to quarantines. We believe our results may also contribute to the theory of interacting particles.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Diffusion processes in networks permeate several distinct domains. The abstraction landscape encompass from electric pulses across neural networks in the brain and fluids passing through porous materials, to consensus emerging in distributed systems, fires spreading through tropical forests, and cascading failures occurring in large-scale power grids. The modeling of such phenomena is based on the network epidemics framework, wherein *epidemic* refers to any iterative dynamic by which viruses, ideas, rumors, failures, or any other propagators can *infect* neighbors of their carriers.

Though an appropriate modeling must take into account each targeted domain's specificity, the general network epidemic framework typically embeds three core features:

1. *Contact pattern.* Traditionally, a *network of contacts*—upon which the propagation spreads—often corresponding to a static, undirected graph where nodes represent individuals and an edge linking two nodes means that these actively interact with one another. In this particular case, the set of individuals one interacts with remains the same throughout the epidemic process. Many other abstractions have been proposed in order to capture more complex contact patterns. For instance, one may consider continuous mobility models in a unit-square region, wherein two nodes interact (edge appearance) whenever their (euclidean) distance lies below some previously fixed threshold. Note that in this case the network of contacts is dynamic (since every possible edge may appear and disappear several times) but the mobility is *unstructured* since an individual is free to move in any direction. A more recent approach, which combines both dynamic contacts and structured mobility, is the *agent-based* model (ABM). Therein, individuals are represented by agents that move on a network. Nodes represent sites and an edge linking two nodes indicates a possible transit along them. Agents thus "walk" (move) from one site to another

obeying topological (structural) constraints.

2. *Propagation laws.* The interactions among individuals are key for an epidemic to unfold. Proper modeling, however, demands further design decisions: once a contact is established between an infectious individual and a healthy one, does it suffice for the infection to be passed along? Does the exposure time matter? For how long does an infected individual remain in such a state? Does an individual's contact pattern change after he becomes infected? If recovered, will the individual become again subject to new infections? Appropriate answers clearly depend on the application. However, the network epidemics framework allows us to consider many of these as adjustable parameters within a *compartmental* setup. These compartments—first proposed by Ross and Kermack & McKendrick [2, 3]—partition individuals across *epidemic states*. For instance, outbreaks wherein an individual's state changes solely from *susceptible* (S) to *infected* (I) are called SI epidemics. If a previously infected individual eventually becomes again susceptible, then dynamics are SIS. Conversely, to acquire immunity past the infection leads one to the *recovered* (R) state, which in turn characterizes an SIR epidemic. Once suitable compartments are defined, the remaining parameters (e.g. *pathogen's transmissibility* and *recovery rate*) must still be determined.

3. *The set of early spreaders*—named epidemic *seeds*—whose infection is generally accounted an exogenous event. This special set plays a central role in the epidemic process, as these propagation starters often influence the size and the dissemination speed of an epidemic outbreak.

In this thesis our investigation concentrates on the two first aforementioned aspects of network epidemic models. In particular, the main focus resides in characterizing *protection effects* (to be soon described), which are closely related to aspects 1 and 2 above.

We remark on two major works derived from each problem considered herein, both later described in detail.

- *Characterizing Protection Effects on Network Epidemics driven by Random Walks* [4], received the *Best Paper Award* at *WPerformance 2020*, a satellite workshop from Brazil's main academic event on Computing, the *Brazilian Society for Computing's Congress* (CSBC).

- *Biased Random Walks as Protection Effects on Agent-based Network Epidemics.* A paper—about to be submitted to IEEE Transactions on Network Science and Engineering (TNSE)—consolidating the present research, and including the most recent findings, which will also be covered in this text.

1.1 Protection Effects

Understanding how epidemics either evolve and die out is increasingly pursued within various disciplines, for reasons such as to avoid or reduce potentially catastrophic impacts across society on its many spheres. The arguably most striking example is the recent COVID-19 pandemic, against which an unprecedented scientific engagement came to be daily witnessed worldwide. Efforts ranged from multiple level of quarantine and conception and launch of effective vaccines to the understanding of spreading patterns [5], socio-demographic responsiveness [6, 7], general psychological trauma [8] and even hesitancy level on taking vaccines [9], to name a few.

In this context, *protection effects* (PrE), aka *disease-behavior dynamics* [10–12], denote the set of measures individuals take to avoid contagion (such as to wear masks and avoid certain locations), once aware that an epidemic unfolds nearby. PrE fundamentally differ from *interventions*, which refer to epidemic containment policies driven by governments: whereas interventions are carried out collective-to-individual (e.g. closure of schools, shops, restaurants, and flight cancellations [13, 14]), PrE arise as a behavioral product of individual risk awareness; an individual-to-collective process. Indeed, intervention is typically encoded as a reduction in the number of contacts per unit time, possibly leading the epidemic to die out. PrE, in turn, are generally translated into a saturation level for the infection rate.

Recent studies have investigated how PrE may impact the course of an epidemic as predicted by traditional models. Most, however, elaborate over static, homogeneous mixing premises and lack networks. While such approaches prove resourceful at capturing and elucidating many key aspects towards disease outbreaks, these also typically neglect, totally or partially, the structure of the underlying network of contacts and its time-varying nature.

This thesis proposes a simple agent-based model with PrE for continuous-time SIS epidemics on non-regular, degree-uncorrelated networks. Protective behavior comes through biasing local agents movements towards safe sites, in an epidemic state-dependent fashion: susceptible (S) agents avoid locations hosting infected (I) agents and vice-versa, with a single parameter to represent aversion strength. Propagation occurs through direct contact between S- and I- agents at a given node and depends on the total exposition time (details are first provided in Chapter 3 and then extended in Chapter 5).

Key factors to modeling epidemics through agent-based schemes include their spatio-temporal expressiveness. Indeed, to take networks for topological structures mobile agents transit into is a natural way of capturing real world mobility patterns [15]. Models of this flavor have been recently considered [15–18]. To the best

of our knowledge, however, non-regular networks and PrE have hitherto never been theoretically investigated under such schemes; a gap filled by this work.

1.2 Contributions

The main contributions of this thesis are summarized as follows.

- For sparse regimes in degree-uncorrelated networks, degree-block approximation is leveraged in order to obtain an accurate epidemic predictor that explicitly embodies structural and protective information into a system of ordinary differential equation (ODE) orders of magnitude smaller than the network size. These results are supported by numerical evaluation of the ODE system and by simulation of the the epidemic network model;
- The first theoretical evidence of a key structural impact on agent-based epidemic models: larger *degree heterogeneity* boosts the spreading;
- Remarkably, we find a special regime under which epidemics on networks with *arbitrary degree distribution* can be accurately predicted in $\Theta(1)$ (constant running time), not by a system but by a *single* ODE equation;
- Protection thresholds for disease-free steady states under both constant and arbitrarily large rates for walk and transmissibility;
- Interestingly, we find a law through which a simple random walk (SRW) can be conveniently combined with the degree function in order to yield *any desired moment of a network's degree distribution.*
- A numerical study on *dense regimes.* Surprisingly, rich, oscillating dynamics arise as a result of the combination of strong PrE versus highly heterogeneous topology. This offers new insights on why we frequently observe real-world *seasonal* epidemics that persist over only a small fraction of the population;
- Design and implementation of a publicly available network epidemic simulator whose numerical results validate the theoretical analysis on the proposed models.

1.2.1 Prior Contributions

Throughout the first two years of his doctorate, the author has also investigated and published results on another epidemic related problem, for which we shall now provide a small synthesis, without going into detail. This decision takes into account

the fact that such an investigation (to be briefly presented below), notwithstanding its relevance, does not account for the main focus of this text nor can it be directly integrated in a complementary fashion to the main work presented herein. A high-level description of this prior work is presented below.

Budgeted Influence Maximization (BIM)

The influence maximization (IM) problem has been broadly investigated since the seminal work of Kempe et al. [19]. This pioneering paper provides a framework for the general problem, proving its NP-hardness and providing an approximate, polynomial-time greedy algorithm with performance guarantees (constant factor from optimal). Their algorithm is based on submodular objective functions, which is shown to be the case for some diffusion models. However, its high running time has led to a myriad of approaches to tackle the problem with more efficient algorithms [19–30]. Indeed, various prior works have focused on designing heuristics to determine good seeds, exploring structural features of the network as well as features associated with nodes (e.g., labels). For example, computationally-inexpensive heuristics based on node degree [22], particle swarm optimization [31], and node homophily [24] have all been considered.

Heuristics based on k-core decomposition [32] have also been explored [21], showing a correlation between influential spreaders and highly connected regions of the network. This idea has been explored by various subsequent works that also adapt and augment such metric with node rankings [28], communities [26], disjoint paths [25], and local neighborhoods [27].

The IM problem has also been investigated under diffusion models fundamentally different from the widely adopted Independent Cascades (IC) and Linear Threshold (LT). For instance, Ugander et al [33] propose the *structural diversity* model, further investigated by Wenzheng et al [34].

However, all these prior works implicitly assume that nodes have identical costs, since the constraint to start a propagation is simply the number of seeds.

There are also recent works that have investigated network seeding where node costs are not fixed (over time) nor identical across the network. For example, Leskovec et al. [35] propose strategies for placing sensors on a network to more quickly detect a diffusion. Arthur et al. [20] propose strategies to price products and provide cash-back (discount) to nodes in the network to induce recommendations to their neighbors. Miyanchi et al. [36] formulate an optimization problem wherein a fixed budget is allocated to a bipartite network of marketing channels and customers with variable node costs (no diffusion considered). None of these works specifically addresses the BIM problem.

However, BIM has more recently been formulated and investigated by Nguyen

and Zheng [37]. The authors depart from the framework introduced by Kempe et al. [19] and tackle the problem using the IC model. They establish a submodular, cost-normalized objective function, from which they determine a greedy algorithm—here called GR—with approximation guarantees up to a constant factor.

Other works have also investigated the BIM problem [38–40]. Han et al. [38] tackle BIM with a heuristic combining two seeding strategies; one based on node influence, and the other on node cost.

More recently, Nguyen et al. have formulated a more general problem, *Cost-aware Targeted Viral Marketing* (CTVM) [39], briefly described as follows. Beyond arbitrary selecting cost, each node v also provides an arbitrary benefit $b(v)$ for being activated. The goal is thus to maximize not the influence spread but the total benefit provided by the final active set. Besides CTVM their algorithm, named BCT, also tackles either the classical IM and the BIM problems. The latter—which is the scope of this work—corresponds to the case where, given a network $G = (V, E)$ and a constant $C \in \mathbb{R}_+^*$, $b(v) = C, \forall v \in V$. They show that, for IM, BCT significantly outperforms state-of-the-art algorithms such as TIM/TIM+ [41] and IMM [42] in terms of running time, with equal performance in what regards the spread of influence. Also, when considering arbitrary selecting costs, they report BCT outperforms all above-mentioned algorithms, including GR, in terms of conceiving a seed set that yields a final active set with larger overall benefit. For BIM, however, they report GR performs better than BCT in terms of total influenced population. Last, Souza et al. [40] characterize the performance of simple and traditional seeding strategies to solve the BIM problem, motivating the need for more clever strategies.

Despite addressing the BIM problem, these prior works have the following limitations. The theoretical result of Nguyen and Zheng [37] assumes that the initial budget is larger than the cost of any node. Moreover, their numerical evaluation uniformly assigns random costs to nodes, from a small range (less than a factor of 10). Similarly, Han et al. [38] and Nguyen et al.[39] assume that the initial budget is larger than the cost of any node, and their numerical evaluation considers that cost and node centrality are linearly related. These assumptions fall short of capturing more general pricing practices, such as those adopted by celebrities (nodes) for promoting viral marketing in online social networks [43–45]. In particular, marketing campaigns may not have sufficient budget to hire even one of those more expensive individuals.

The above context paved the way to the author's first major contribution made along his Doctorate's degree pursuit. The full text has been published in the journal *Information Sciences* [4] in 2020. Therein, we propose a flexible node cost model that strictly depends on the network structure and allows for an arbitrary range

of values, without making assumptions on the available budget. Efficient seeding strategies for BIM were investigated by the author under a correlated node cost model (i.e., higher centrality, higher cost) and the linear threshold model. The main contributions of the aforementioned publication are as follows:

- We propose a seeding strategy, called *Node Surround*, which consists of targeting the cheapest neighbors of central, expensive (or even cost-prohibitive) nodes, leveraging their higher spreading potential at much lower costs when compared to that of their direct seeding. We show that, as the network threshold increases, this approach outperforms state-of-the-art BIM strategies.

- We show how the classical *fraction of activated nodes*, a broadly adopted metric, may lead to misinterpretations with respect to the effectiveness of a strategy. Opposite to the unit-cost IM, different BIM strategies with the same initial budget may still yield seed sets of *very different sizes*, ranging from few key-nodes up to a large fraction of the network. By considering solely their *diffusion power* (a metric we have proposed), we capture the real benefit (activated non-seeds) of an investment (budget). Diffusion power (DP) is a fundamental metric to properly assess BIM strategies. It embeds the *Outward Influence* [46] concept—originally proposed for IM—to tackle BIM. Indeed, we show that to ignore the seeds (paid influencers) when measuring a strategy's performance eliminates potentially large assessment distortions.

- We propose a flexible, single-parameter model for node cost, which correlates cost and network centrality. To the best of our knowledge, this is the first model to admit non-linear relations between cost and local structure. It also captures the common real-world scenario wherein node-costs across a network may differ from one another by *many orders of magnitude*, thus being more relevant in practice.

We refer the reader to [4] for a complete description on the problem. As for this thesis, no further details on this specific work is to be provided hereafter. Instead, in what follows we switch back to the original agent-based epidemic problem firstly introduced, and elaborate on such an investigation front.

1.3 Text Organization

The remainder of this thesis is organized as follows. In Chapter 2 we provide background on the problem and contextualize it through a discussion of related works. Chapter 3 describes our most basic predictor, designed for complete graphs, with theoretical analysis and simulation results. Then, Chapter 4 substantially expands

the previous model in order to accommodate more general network structures. Chapter 5 describes how to embed PrE into the model formulated in the previous chapter, thus obtaining our final model. A more involved theoretical analysis is carried out in this chapter, and a series of simulation results show the accuracy of our model. Last, we conclude the thesis on Chapter 6.

Chapter 2

Background and Related Work

2.1 Epidemics and models

Modeling propagation dynamics that unfold upon direct contact between individuals, specially within the field of mathematical epidemiology, has enormously benefited from the compartmental approaches proposed a century ago by Ross [2] and Kermack & McKendrick (KMK) [3]. Their work provided fundamental insights on the dynamics of susceptible-infected-susceptible (SIS) and susceptible-infected-recovered (SIR) epidemics, respectively, and inspired many models to be crafted on top of other compartments, such as SI and SIRS [47].

The general framework behind such models is to consider the entire population as partitioned into non overlapping epidemic states—the so-called *compartments*—and then establish (i) what is the set of early spreaders, and (ii) what are the epidemic laws that govern how individuals migrate between compartments once an outbreak has started. Once these two features are established, prediction of how infection propagation unfolds is then typically obtained by numerically solving an ODE system that incorporates these properties.

In what follows we briefly present some of the most widely studied classic models. The interested reader should refer to [47] for more details on each of them.

2.1.1 SI

This is the simplest classic epidemic model. Its name comes after *Susceptible* (S) and *Infected* (I), the only two compartments it considers. Here, the only possible epidemic transition is from S to I. Once infected, an individual remains so until the end of the epidemic. Clearly, any non empty set of early spreaders, denoted I_0, leads to the entire population eventually become infected.

This model can be segmented into three core phases: (i) sub-critical, wherein just a few individuals carry the disease, thus making the propagation unfold slowly; (ii)

critical, which refers to the onset of a far-reaching, exponentially fast propagation, and (iii) hypercritical, when the spreading finally reaches a saturation level after having already infected a huge fraction of population. Those remaining susceptible individuals are again infected at a slow rate when compared to the critical phase.

Denoting by τ the transmissibility rate, SI dynamics over a population of $K = S + I$ individuals is modeled as

$$\begin{cases} \dfrac{dI}{dt} = \tau \dfrac{SI}{K} \\ \\ \dfrac{dS}{dt} = -\tau \dfrac{SI}{K}, \end{cases} \tag{2.1}$$

or, in terms of densities $s = S/K$ and $i = I/K$ (thus $s + i = 1$),

$$\begin{cases} \dfrac{di}{dt} = \tau si \\ \\ \dfrac{ds}{dt} = -\tau si, \end{cases} \tag{2.2}$$

but since population is partitioned, we may notate $s = 1 - i$ and reduce the whole dynamics to a single equation, which reads

$$\frac{di}{dt} = \tau i(1 - i). \tag{2.3}$$

Interestingly, (2.3) does actually correspond to the well-known *logistic growth equation*, broadly employed across various disciplines.

2.1.2 SIR

Although not the most trivial, this is in fact one of the first compartmental models proposed, formulated a century ago by KMK. Here, in addition to S and I, a third compartment, *Recovered* (R) (also known as *Removed*), embraces individuals that either obtain immunity some time after acquiring the disease, or die as a consequence of it. The only epidemic transitions such a model admits is from S to I and from I to R. Contrary to the SI model, here it is clear that the entire population will eventually integrate compartments R (predominantly) and S (residually), with no remaining infectives.

Denoting by τ and γ the rates for transmissibility and recovery, respectively, SIR

dynamics is modeled as

$$\begin{cases} \dfrac{dI}{dt} = \tau \dfrac{SI}{K} - \gamma I \\ \\ \dfrac{dS}{dt} = -\tau \dfrac{SI}{K} \\ \\ \dfrac{dR}{dt} = \gamma I, \end{cases} \tag{2.4}$$

which in density terms reads

$$\begin{cases} \dfrac{di}{dt} = \tau si - \gamma i \\ \\ \dfrac{ds}{dt} = -\tau si \\ \\ \dfrac{dr}{dt} = \gamma i. \end{cases} \tag{2.5}$$

2.1.3 SIS

It is specially worth mentioning this classical approach since the present thesis is based on its fundamental concepts. SIS stands for Susceptible-Infected-Susceptible, and its name reflects the fact that, once infected, an individual can eventually become susceptible again. Therefore, one same individual may contract the same disease multiple times under such a model. Whereas SI and SIR have stationary regimes where the I compartment embraces, respectively, all individuals or virtually none of them, SIS presents a stationary regime where only a fraction of the population remains infected. This portion can be any value between 0 and 1.

Denoting by τ and γ the rates for transmissibility and recovery, respectively (as in the case of SIR), dynamics are given by

$$\begin{cases} \dfrac{dI}{dt} = \tau \dfrac{SI}{K} - \gamma I \\ \\ \dfrac{dS}{dt} = -\tau \dfrac{SI}{K} + \gamma I. \end{cases} \tag{2.6}$$

Its representation in terms of density reads

$$\begin{cases} \dfrac{di}{dt} = \tau si - \gamma i \\ \\ \dfrac{ds}{dt} = -\tau si + \gamma i. \end{cases} \tag{2.7}$$

At this point it is clear that the choice on the appropriate model depends on the actual pathogen under consideration. For instance, diseases that once contracted lead their hosts to acquire lifetime immunity are better modeled via SIR (Susceptible, Infected, Recovered). If, conversely, the same disease may come to be contracted again right after the individual gets healed, then SIS is more suitable, and so on.

Unfortunately, these models are known to have various limitations. In particular, they implicitly assume *homogeneous mixing* and *infinitesimal population*. The former refers to the fact that any such model assumes that the population is *fully mixed* i.e., that at any point in time, any individual from any given compartment directly interacts with any other individual from any other compartment, without topological constraints. The latter refers to the fact that these models relax the real-world restriction on integer values since arbitrarily small *fragments of individuals* are accepted, just like a fluid. Such a relaxation however is arguably not a major problem, since the main purpose behind these models is not to predict an epidemic's exact outcome, but its *average progress* over time. Homogeneous mixing, on the other hand, configures a major unrealistic drawback, as no real-world epidemic behaves like this.

The next section presents the natural evolution of the models hitherto discussed. Instead of a fully mixed population, it encodes a more natural *network of contacts*, through which interactions between individuals obey to structural constraints, as detailed in what follows.

2.2 Classic Network Epidemics

In order to cope with the strong analytical limitation imposed by the homogeneous mixing premise inherent to all classic models, Pastor-Satorras & Vespignani [48] proposed a framework for predicting more complex, networked contact patterns on SIS epidemics, which was later expanded to other compartmental setups, such as SI and SIR [1]. Their celebrated work provides an extension for the classic predictive models seen in the previous section, by accommodating the fact that individuals in society are generally very heterogeneous in what regards their number of contacts and the actors within their social circles, i.e. who interacts with whom.

Prior to presenting the model, we must describe the *purpose* behind the use of networks in spreading processes. Indeed, networks are long known to be key ingredients for the study of dynamical processes involving the spread of propagators, in particular among individuals. For instance, these have been made use of in sociology 40 years ago to model influence propagation across society [32]. In general, two main purposes support their adoption:

- Modeling the epidemic itself. In this case, simulations over the obtained setup

are conducted so that useful observations and intuition can be extracted out of it;

- Modeling a *predictor* on the epidemic's average outcome. This is the huge leap of [48], as it provisions a framework for encoding structural information within ODE's, so that a network epidemic's average outcome can be more appropriately predicted by fluid models.

We shall now briefly describe this extension, which has long been instrumental to the field of Network Epidemics. The core aspect of [48] is to encode interactions that, rather than fully-mixed, adhere to a *network of contacts* instead. That is, individuals are now encoded as nodes in a network, and possible interactions are expressed by edges between pairs of nodes. An edge linking two nodes mean that these keep contact with one another, and are therefore *neighbors* within the network. In this case, a direct channel for disease transmission—the edge—exists between them.

Let the *node degree* $d(v)$ of a node v be the number of edges incident to v in an undirected network $G = (V, E)$ i.e., $d(v) = \sum_{u \in V} \mathbb{I}((v, u) \in E)$, where $\mathbb{I}(\cdot)$ is the indicator function. Pastor-Satorras & Vespignani [48] assume that the network is *degree uncorrelated* i.e., that for any edge uniformly chosen at random, the degree of one of its nodes is independent of the degree of the other node. This property is often measured in practice through a metric called *assortativity coefficient*, firstly proposed by Newman [49]. In short, this metric assigns to a network some number in the interval $[-1, 1]$, where -1 means a fully *dissortative* network, and 1 means a fully *assortative* one. A network is said to be assortative if it exhibits the trend of having nodes of similar degrees connected to each other. Conversely, dissortative networks present connection patterns wherein neighboring nodes tend to have very different degrees, i.e., large degree nodes are more often linked to low degree ones and vice-versa. Therefore, the assumption of a network being degree uncorrelated means that its assortativity coefficient must be zero (or very close to zero).

2.2.1 Degree-block approximation

The next key ingredient in the network contact model—directly derived from the no degree-correlation premise—is the possibility of handling all nodes of the same degree, say b, together, since these are now assumed to be statistically equivalent. This technique is termed *degree-block approximation*. Within this paradigm, every node from a certain block b (i.e., every node v such that $d(v) = b$, $b \in \mathbb{N}$) is equally likely to be the recipient of an edge whose other extreme is a node of degree b'. This degree-block separation scheme is illustrated in Figure 2.1, which is original from [1] and reproduced below:

Figure 2.1: Degree-block approximation (image reproduced from [1]). Nodes not only preserve their epidemic state partitioning (in this case, full circles for I and open circles for S), but can now also be partitioned into degree blocks.

The idea is thus to treat nodes from a same epidemic state *and* same block under one same equation. In what follows we show the form of an SIS epidemic model under this classical network epidemic approach.

2.2.2 SIS epidemics on networks

The basic SIS model for networks is given (in terms of density) by the following system (which applies for each degree block b):

$$\frac{di_b}{dt} = \tau(1 - i_b)b\Theta_b(t) - \gamma i_b. \tag{2.8}$$

(2.8) states that the rate at which the density of infectives from some block b varies with time is governed by a few key quantities. First, note that since every block has its nodes partitioned into S or I, the density for S-nodes is simply $1-i_b$. Also, because we no longer assume homogeneous mixing, there is a limit of b connections through which a degree-b node interacts with the network, plus a *probability* $\Theta_b(t)$ that an S-node becomes infected at the continuous time t. For the interested reader, we refer to [1] for a complete derivation of $\Theta_b(t)$ (which is based on the no degree-correlation assumption). Last, the density of infectives decays with those that recover at rate γ.

Despite its enormous relevance, the classical network epidemics model has two major shortcomings: first, it considers a *static* network of contacts i.e. nodes do always preserve their same connections along with time. Likewise, it also implicitly assumes that nodes do always behave naively when exposed to risk. It is clear, however, that in real-world situations individuals adapt their routines upon temporary

threats underway. Indeed, several models have since then proposed variations for more complex behavior—in the form of PrE and/or interventions—but with only small achievements with respect to agent-based approaches.

Noteworthy, one key difference between the above formulation and the one to be presented in this thesis is that, in our case, a node is not an individual but rather a site. Individuals are encoded by mobile agents, which may co-exist in a same node irrespective of their state. Consequently, even though we leverage degree-block approximation, we cannot assume, for a given block b, that $s_b = 1 - i_b$, hence both populations (S and I) in our model will require explicit equations within each block (more details later).

2.3 Epidemics with Protection Effects

Roughly 50 years after KMK's work, Capasso and Serio [10] provided a generalization of KMK's model, replacing the (fixed) infection rate by an infection function $g(\cdot)$, thus capturing more general interactions between susceptible and infected agents. They show that $g(\cdot)$ can play two different roles, namely intervention or protection effects. Their work is among the pioneers in considering such concepts as key ingredients for predicting an epidemic's outcome.

Disease-behavior dynamics has ever since been investigated from many different perspectives [12, 50]. For example, Hyman and Li [51] consider SIS epidemics of sexually transmitted diseases. The authors formulate a mean-field model that segments population into risk-level groups, and conclude that behavioral changes (such as reducing contacts and partner formations) may decrease the infection level. Tchuenche et al. [52] turn attention to the influence of local media on population's adaptive behavior to an ongoing outbreak, and conclude that media coverage does not necessarily help to promote epidemic containment. An intervened SIRS epidemic (i.e. a SIRS epidemic with intervention forces) [53] and a SIS epidemic with PrE (induced by media coverage) [54] were both investigated under Stochastic Differential Equations models. In both cases, it is shown that outbreaks otherwise endemic may still die out due to large random fluctuations.

Risk awareness has also been analyzed in the context of multiplex networks. For example, Granell et al. [55] consider a network of physical interactions—through which an SIS epidemic spreads—coupled to a (virtual) social network wherein the same actors disseminate awareness in a fashion similar to an SIS epidemic. The authors show that the propagation of awareness may delay or even preclude epidemics that would otherwise yield large outbreaks. Mao and Yang [56] propose a framework for modeling PrE in multiplex networks and draw particular attention to the fact that real world infection rates may be significantly larger than those predicted by

models not encoding PrE. Indeed, the rate as estimated by these models consider that the affected population behaves naively with respect to the epidemic, when people do actually adapt and take protective measures throughout the process.

When the time-varying nature of real world contact patterns is taken into account, outbreaks may yield dynamics that traditional models fail to capture [57, 58]. Within this paradigm, Robinson et al. [59] provide evidence that Great Britain's restrictions on cattle movement (between animal holdings) in order to avoid epidemics had gradually lost efficiency. Remarkably, the pointed reason is the self-organizing network induced by behavioral changes from farmers, who came to intensify cattle movement across the network's giant strong component. Particularly intriguing, Zhou et al. [60] provide insights on the protective dynamics that could possibly explain seasonal epidemics. Finally, Yang et al. [61] study the impacts of *emigration* as a protective maneuver. In their model, mobile agents are free to walk in any direction within a square region and move only upon imminent risk, choosing a new location uniformly at random. They conclude that protective actions performed sufficiently early may avoid endemic steady states.

2.4 Agent-based Network Epidemics

None of the prior works addresses PrE in epidemics where agents move within a network and locally attempt to avoid contact with one another. On the other hand, agent-based models have predominantly focused on fitting problems i.e., the establishment of models capable of accurately explaining past epidemics up from real-world data, so these calibrated models can then be applied to estimate future outbreaks. Despite their critical social relevance, not only do these models require the availability of massive amounts of data [62] but also often encompass so many parameters it becomes hard to determine the extent to which each of them influences the observed outcome.

This thesis aims at understanding the fundamental impacts of the interplay between arguably the most central features in agent-based network epidemics: network structure and mobility patterns, a goal of both theoretical and practical interest. Indeed, such an understanding will allow for better decision making towards real-world outbreaks. On the other hand, among the various ways through which one may encode epidemic processes, agent-based models (ABM) in networked environments are known to be best for fine-grained representations of complex interactions between individuals at a micro-scale [63]. Despite, these are also the most challenging in terms of providing theoretical insights [64]. Even to date, theory in the field is still exiguous, mostly built on top of very specific setups (e.g. lattices or cliques). This fact highly contrasts with the popularity such models have increasingly been

enjoying through the last two decades [63].

Frameworks for ABM that cope with more general network structures and non trivial mobility while providing basic statistical mechanics, threshold values and analytical tools are thus eagerly welcomed. This work is a step in such a direction. True, a broader analytical view on the extent at which major features of such processes influence their associated dynamics is key for designing more complex, predictive models.

Chapter 3

Protection Effects on Complete Networks

This chapter fully describes the protective mobility scheme briefly introduced in Chapter 1 when the underlying network is a complete graph (i.e. a clique). First, we provide a detailed description of SIS epidemics from a mobile-agent perspective in Section 3.1. The protective behaviour is then described and coupled to the model in Section 3.4 and analysed in Section 3.6. Simulation results that support our theoretical findings (in this chapter, at the scope of complete graphs) are presented and discussed in Section 3.7.

3.1 SIS epidemics with mobile-agents

We shall now describe the coupling of SIS epidemics with a mobile-agent environment, as considered in this work. The following analysis—which still does not consider PrE—follows closely the work of Ibrahim [17] but has been modified in order to capture infections that depend on the exposure time.

Table 3.1 lists the main terms and parameters to be used throughout this chapter.

Mobility and contact pattern. Consider an undirected network $G = (V, E)$ with node and edge set given by V and E, respectively, where $n = |V|$ denotes its size. Consider a set K of $k = |K|$ agents, and let $v_j(t), j = 1, \ldots, k$ denote the location of agent j in time $t \geqslant 0$. Note that $v_j(t) \in V$ as agents can only be found in network nodes. At time zero, the location of an agent is chosen uniformly at random from V. Agents move according to continuous time random walks, where the residence time in any given node is exponentially distributed with rate $\lambda > 0$ (the walk rate). Once the agent has to move, it chooses its next node uniformly at random from the neighboring nodes. Such transitions are assumed to occur instantaneously. Any two agents j and l are assumed to be in contact with one another *iff* both of them are

Table 3.1: Symbols & terminologies

S-agents	Susceptible agents.
I-agents	Infected agents.
SI-contact	Contact (i.e. encounter) between an S-agent and an I-agent.
$G = (V, E)$	Undirected network wherein mobile-agents walk.
n	Network size, such that $n = \|V\|$.
$N(v)$	Set of neighbors of node v.
K, k	Set of agents, such that $k = \|K\|$.
$S(t), I(t)$	Set of S/I-agents at time t. $S(t) \cup I(t) = K$, $S(t) \cap I(t) = \varnothing$.
S, I	$\|S(t)\|$ and $\|I(t)\|$, respectively, thus $S + I = k$
s, i	S/k and I/k, respectively, hence $s + i = 1$.
i_0	Fraction of initially-infected agents, i.e. when $t = 0$.
τ	Disease transmissibility.
λ, β, γ	Walk rate, infection rate and recovery rate, respectively.
σ	Infection probability, such that $\sigma = \tau/(2\lambda + \tau)$.
α	SI-contact rate.
C	A constant, such that $C = (\tau k \lambda)/((2\lambda + \tau)n)$.
w_s	S-agent's tolerance to SI-contacts, $0 < w_s < 1$.
w_i	I-agent's tolerance to SI-contacts, $0 < w_i < 1$.

located at the same node, $v_j(t) = v_l(t)$. Note that this leads to a dynamic *network of contacts* that is time-varying and can be characterized by a collection of isolated cliques.

Epidemic state and infection. Besides residing in a node, every agent has an epidemic state, denoted by "S" (susceptible) or "I" (infected). Let $c_j(t) \in$ { "S","I" } denote the epidemic state of agent j at time $t \geqslant 0$. Also, let $S(t) = \{j \in K | c_j(t) =$"S"$\}$ and $I(t) = \{j \in K | c_j(t) =$"I"$\}$ denote the set of susceptible and infected agents at time t, respectively. Note that $S(t) \cup I(t) = K$ and $S(t) \cap I(t) = \varnothing$ for all t. Disease spreads through direct contact between an S-agent and an I-agent, with infection probability proportional to the duration of such a contact. The decision on whether or not an S-agent becomes infected is taken the moment it leaves its current node, as follows. Let t_e denote the total time an S-agent a remains exposed to one or more I-agents while residing in some node. Considering an exponential random variable Y with parameter $\tau > 0$, the probability that a becomes infected is simply $P[Y < t_e]$. New infections thus depend not only on t_e but also on the pathogen's *transmissibility rate* τ, such that within some fixed exposure interval an agent is more likely to become infected as τ increases.

Recovery. Note that agents can only become infected when taking a step, moving to some node. Once infected, an individual remains so for a certain time window, recovering right after. During such a period, however, the walker may infect others. The elapsed time until an agent recovers is assumed to be exponentially distributed with rate $\gamma > 0$—the *recovery rate*—and independent of any other events.

As with other proposed random walk models [15, 17], a *sparse* scenario is assumed, where the number of network nodes is much larger than the number of agents. In addition to being a good approximation for real-world sparse cases, this also allows the model to be simplified by assuming that only pairwise encounters occur, as the probability of having three or more agents in a same node becomes negligible. This assumption is fundamental in the following analysis.

3.2 Problem Statement I

Our first goal is to establish a fluid model that predicts the average behavior of epidemics governed by the aforementioned laws. Later in Section 3.4 we extend the problem at hand by including protective behavior, manifested as biased random walks.

3.3 Modeling epidemic dynamics

The following deterministic model predicts the expected epidemic dynamics under the assumption that the network is a complete graph (i.e., all possible edges are present). As with the classical models, it relies on ODEs to capture the change in the population of S-agents and I-agents. In order to simplify notation, let $I = |I(t)|$ (resp. $S = |S(t)|$) denote the number of infected (resp. susceptible) agents at time t. Then, the rate dI/dt at which the infected population changes over time depends on five different quantities, as follows.

1. The rate at which two given agents meet. Since G is a complete graph, this rate can easily be shown to be $2\lambda/n$, since their joint walk rate is 2λ and in one step they can reach any node $v \in V$ with probability $1/n$.

2. The number of possible encounters among the agents. Since the sparse regime is assumed, only pairwise encounters are possible (with high probability), and thus there are a total of $\binom{k}{2}$ possible encounters;

3. The probability of an SI-contact, given by the number of possible SI-contacts at time t divided by the total number of possible contacts, i.e. $SI/\binom{k}{2}$;

4. The probability of infection given an SI-contact, denoted by σ. This depends on the transmissibility τ and the total time the S-agent remained exposed, i.e. the duration of the SI-contact. Since rates for both transmission and walk are governed by exponentials, it can be shown (Appendix A) that

$$\sigma = \frac{\tau}{2\lambda + \tau}. \tag{3.1}$$

5. The total rate at which I-agents recover and become S-agents is simply given by γI, by independence among the agents in the recovery process.

Thus, the infected population dynamics is given by:

$$\frac{dI}{dt} = \binom{k}{2}\frac{2\lambda\sigma SI}{n\binom{k}{2}} - \gamma I = \frac{2\lambda\tau}{n(2\lambda+\tau)}SI - \gamma I, \tag{3.2}$$

which in terms of $i = I/k$ (and noting that $S = sk$), becomes

$$\frac{di}{dt} = \frac{2\lambda}{n}\frac{\tau}{2\lambda+\tau}ksi - \gamma i = \beta si - \gamma i, \tag{3.3}$$

where the rightmost term exhibits the classical SIS form (Eq. 2.7), with $\beta = \frac{2\lambda\tau k}{(2\lambda+\tau)n}$.

3.4 Protection model

We consider the protective behavior induced by how agents elect each next-hop during their random walks. Decisions are made locally, based on the instantaneous information from neighboring nodes only. PrE are thus represented as *biased* random walks: every agent j has its walk biased so that hostile nodes are avoided.

Let w_s and w_i denote S-/I-agent biases to avoid SI-contacts, respectively, with $0 \leqslant w_s, w_i \leqslant 1$. Now, consider the location of an agent j at time t, namely $v = v_j(t) \in V$. When j moves, it no longer chooses among v's neighbors uniformly, but rather avoids hostile nodes with a bias w_x, where x refers to either s or i according to j's current state. In turn, those edges driving j to safe sites are given unitary weights.

Figure 3.1 illustrates this protective scheme, which we formalize as follows. Let t denote the moment at which an agent j walks, and let $H_{u,v,j,t}$ denote the event "node $u \in N(v)$ is hostile to agent j at time t". The probability that j moves to such node u at t is

$$\begin{cases} \dfrac{w_x^{\mathbb{I}(H_{u,v,a,t})}}{W_{v,a,t}} & \text{, case } W_{v,a,t} > 0 \\ \\ \dfrac{1}{d(v)} & \text{, otherwise,} \end{cases} \tag{3.4}$$

where $\mathbb{I}(\cdot)$ is the indicator function (here as the exponent of w_x), and

$$W_{v,j,t} = \sum_{u \in N(v)} w_x^{\mathbb{I}(H_{u,v,j,t})} \tag{3.5}$$

is the required normalizing constant. Agent next-hop probabilities hence depend on the instantaneous configuration of safe/hostile vicinity. Note that, for the particular

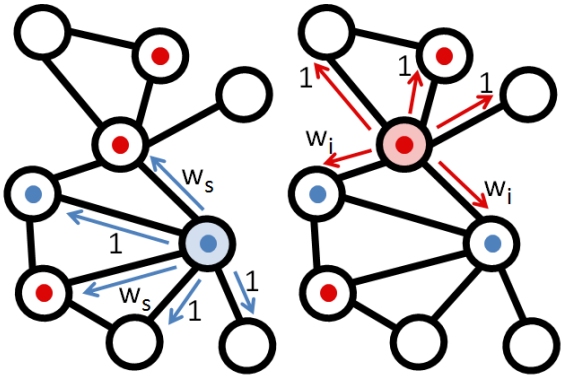


Figure 3.1: Mobility scheme with PrE derived from biased random-walks (self-loops omitted for simplicity). Blue (resp. red) circles represent S-(resp. I-)agents. Weights are dynamically assigned to edges the moment an agent j walks, so that its next-hop tends to safe sites. For instance, should the highlighted S-agent (left) move, and assuming $w_s = 0.4$, its instant next-hop probability assigned to a particular edge would be 0.4/3.8 if it links to a hostile node and 1/3.8 otherwise.

case where $w_s = w_i = 1$—which implies no PrE—dynamics are as in Section 3.3 i.e., agents choose their next location uniformly. Conversely, if $w_s = w_i = 0$ then agents will never step into hostile nodes whenever a safe site is made available. This is the strongest possible PrE, which is likely to end any epidemic.

3.5 Problem Statement II

Our second goal is to establish a fluid model that predicts the average behavior of epidemics governed by the aforementioned laws, where PrE is now encoded. Later in Chapter 4 we further extend the problem at hand by considering more general, degree-uncorrelated networks.

3.6 Theoretical analysis

The proposed PrE model is now coupled with the epidemic model presented in Section 3.3 in order to evaluate key-aspects, namely (i) the SI-contact rate (a fundamental model parameter), (ii) dynamics of infected population, and (iii) the basic reproduction number R_0 (to be later detailed). Again, the network is assumed to be a complete graph.

Since w_s and w_i are block-independent, we may simply consider a single, equivalent protective level w given in terms of their mean, i.e.

$$w = (w_i + w_s)/2. \tag{3.6}$$

Since it allows for a concise and flexible notation, through the rest of this analysis, any occurrence of w must be interpreted as in (3.6).

3.6.1 SI-contact rate

The SI-contact rate now depends on the PrE parameters w_s and w_i. This rate can be represented as the sum of two rates: the rate at which S-agents step into nodes containing I-agents, and vice-versa. These two cases are considered separately, starting with the S-agents, as follows.

Consider an S-agent at the moment it takes a step. The probability it enters a location where an I-agent resides can be computed as follows. Recall that there are I infected agents, each occupying a different location (due to sparsity assumption). Each such location is avoided with bias w_s. There are $n-I$ other locations, each one taken with a bias of 1. Thus, the probability that an S-agent provokes an SI-contact is simply

$$p = \frac{Iw_s}{Iw_s + n - I} = \frac{Iw_s}{I(w_s - 1) + n}. \tag{3.7}$$

Analogously, consider an I-agent at the moment it takes a step. The probability that it enters a location where an S-agent resides can be computed as follows. Recall that there are S susceptible agents, each occupying a different location (due to sparsity assumption). Each such location is avoided with bias w_i. There are $n - S$ other locations, each taken with a bias of 1. Thus, the probability that an I-agent provokes an SI-contact is

$$q = \frac{Sw_i}{S(w_i - 1) + n}. \tag{3.8}$$

The SI-contact rate will thus depend on the number of S/I-agents walking at rate λ each, and their respective SI-contact probabilities, such that

$$\alpha = \lambda Sp + \lambda Iq = \lambda SI \left(\frac{w_s}{I(w_s - 1) + n} + \frac{w_i}{S(w_i - 1) + n} \right). \tag{3.9}$$

Note that α is a fundamental parameter for the model, and is the main modification required in the models presented in Section 3.3.

3.6.2 Evolution of infectives

Under the PrE perspective, the SI-contact rate from Section 3.6.1 must be accommodated into the model. Note, however, that the probability σ of contagion remains the same for an SI-contact, as well as the recovery rate γ. Thus, the change in the

number of I-agents becomes $dI/dt = \alpha\sigma - \gamma I$, i.e.

$$\frac{dI}{dt} = \frac{\tau\lambda}{2\lambda+\tau}\left(\frac{w_s}{I(w_s-1)+n} + \frac{w_i}{S(w_i-1)+n}\right)SI - \gamma I. \tag{3.10}$$

We may rewrite (3.10) in terms of $i = I/k$ and s, noting that $S = sk$. Thus

$$\frac{di}{dt} = \frac{\tau\lambda k}{2\lambda+\tau}\left(\frac{w_s}{ik(w_s-1)+n} + \frac{w_i}{sk(w_i-1)+n}\right)si - \gamma i \tag{3.11}$$

wherein the infection rate β now becomes dependent on i, such that

$$\beta(i) = \frac{\tau\lambda k}{2\lambda+\tau}\left(\frac{w_s}{ik(w_s-1)+n} + \frac{w_i}{sk(w_i-1)+n}\right) \tag{3.12}$$

and hence $di/dt = \beta(i)si - \gamma i$. Note, however, that β's dependence on i is actually negligible. Indeed, the premise of sparsity imposes $k \ll n$ and, consequently, $ik(w_s - 1) + n \approx n$; likewise, $sk(w_i - 1) + n \approx n$. Therefore, a good approximation for (3.12) is

$$\beta = \frac{\tau\lambda k(w_s + w_i)}{(2\lambda+\tau)n}, \tag{3.13}$$

and for a general w (see (3.6)), (3.13) becomes

$$\beta = \frac{2\tau\lambda kw}{(2\lambda+\tau)n}. \tag{3.14}$$

Finally, rewriting (3.11) in terms of (3.13) gives

$$\frac{di}{dt} = \frac{\tau\lambda k(w_s + w_i)}{(2\lambda+\tau)n}si - \gamma i\,, \tag{3.15}$$

whose general counterpart (3.14) is

$$\frac{di}{dt} = \frac{2\tau\lambda kw}{(2\lambda+\tau)n}si - \gamma i\,. \tag{3.16}$$

3.6.3 Reproduction number

The basic reproduction number $R_0 = \beta/\gamma$ is a classical metric which considers the pathogen's spreading potential, and is used to indicate whether the epidemic will die out shortly ($R_0 < 1$) or long-last among population ($R_0 > 1$) [47]. Within the context of PrE, a natural question arises: is there a range of values for w_s and w_i as a function of other model parameters that ensures $R_0 < 1$? In order to answer this question, let us first consider a constant C defined as

$$C = \frac{\tau k\lambda}{(2\lambda+\tau)n} \tag{3.17}$$

so that the infection force β (3.14) may be rewritten as $\beta = 2Cw$. Trivially, because $R_0 = \beta/\gamma$ it follows that $2C/\gamma < 1 \implies R_0 < 1$ for any $\{w_s, w_i\}$, i.e. the epidemic is led to extinction even in the absence of protective efforts. Conversely, if $2C/\gamma \geqslant 1$, then

$$R_0 < 1 \iff \frac{\beta}{\gamma} < 1 \iff \frac{2Cw}{\gamma} < 1 \iff w < \frac{\gamma}{2C}. \tag{3.18}$$

(3.18) exposes that both w_s and w_i, alone, deliver limited protection in the case the other assumes a fixed value. For instance, if infected agents induce no protection ($w_i = 1$) then $R_0 < 1 \implies w_s < (\gamma - C)/C$, which is not possible when $\gamma - C < 0$ since $0 \leqslant w_s \leqslant 1$. On the other hand, it also shows that there always exists some $w > 0$ that forces $R_0 < 1$, i.e. the S-/I-agents joint engagement may positively prevent endemic steady-states. Note, however, that the value for w to satisfy (3.18) depends both on the walk rate λ and the transmissibility rate τ. A particularly interesting question is thus whether there exists a regime for w under which $R_0 < 1$ *even in the case* λ (resp. τ) *is arbitrarily large*, for a fixed τ (resp. λ). This regime can be identified by considering the asymptotic behavior of λ and τ, one at a time, on the infection rate β (3.14). Indeed,

$$\lim_{\lambda \to \infty} \beta = \lim_{\lambda \to \infty} \frac{2\tau\lambda kw}{(2\lambda + \tau)n} = \frac{\tau kw}{n} = \beta', \tag{3.19}$$

which yields a basic reproduction number $R_0 = \beta'/\gamma$. Clearly, $\beta' < \gamma \implies R_0 < 1$. Thus,

$$R_0 < 1 \iff w < \frac{\gamma n}{\tau k}. \tag{3.20}$$

Interestingly, any w satisfying Equation 3.20 will manage to extinguish the epidemic *irrespective of the walk rate* λ. Indeed, larger walk rates, on the one hand, increase SI-contacts per unit time; on the other, they reduce S-agents' exposure time as well. In similar fashion, as $\tau \to \infty$ (for a fixed λ), β converges to $\beta' = (2\lambda kw)/n$. Here, $\beta' < \gamma$ is met case $w < (\gamma n)/(2\lambda k)$. Any such w will succeed at imposing short-lived epidemics *irrespective of the transmissibility rate* τ.

3.7 Numerical results

Next we present simulation results that validate the theoretical analysis from Section 3.6. A discrete-event simulator [65] was designed and implemented [4] in order to generate performance metrics concerning the impacts of PrE and other model parameters. All results to follow assume $i_0 = 0.5$ and simulation time limit $T = 10^5$.

Figure 3.2(a) shows the fraction of infected agents over time for different levels of PrE. Each given w yields two different curves in the plot: the numerical simulation from one single run and the model prediction from (3.15). Note that the model

succeeds in capturing the average dynamics for each protection level. Moreover, the value for w decisively changes the epidemic's outcome. In particular, protective level $w = 0.14$ has managed to hinder the epidemic, leading to a disease-free steady state in a short time. Indeed, from the given parameters, and from (3.18), $w = 0.14 < \gamma/2C = 0.1425$, thus implying $R_0 < 1$.

Figure 3.2(b) shows how R_0 varies as a function of w for two different population sizes, namely $k = 400$ and the $5\times$-larger $k = 2000$. Note that R_0 grows linearly for both cases, but has larger slope when k is larger. The point highlighted in green—for which R_0 is slightly below 1—indicates the protective level of $w = 0.14$ shown in Figure 3.2(a).

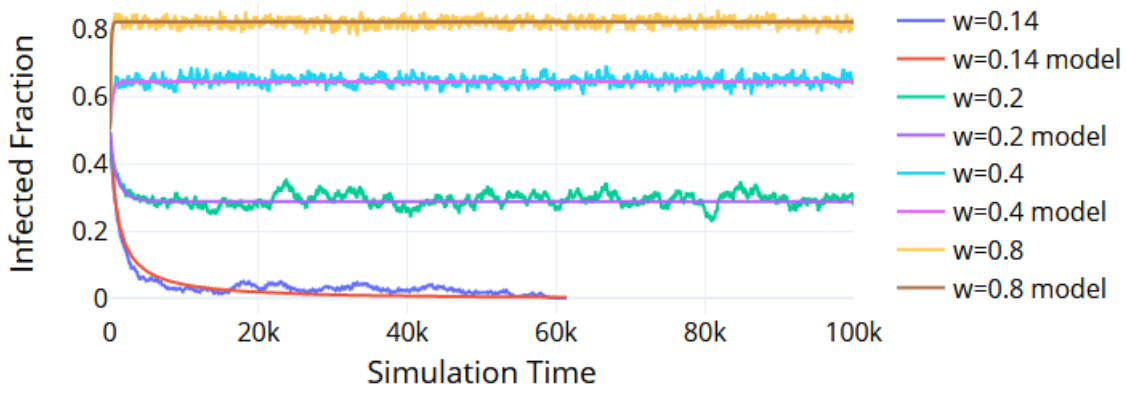
The epidemic's average duration in function of the number k of agents and the walk rate λ are shown in Figures 3.3(a) and 3.3(b), respectively. Each dot from each curve averages over 30 runs. Each curve refers to a different epidemic scenario, in terms of either the network size n and the risk-tolerance level w.

Note that to increase either k or λ leads to a phase transition on the epidemic duration (from ephemeral to long-lasting). More importantly, Figure 3.3(a) illustrates how PrE may drastically increase the epidemic threshold. In particular, by comparing the two epidemics for $n = 10^5$, note that $w = 0.6$ right-shifts the phase transition observed for $w = 1$ (starting around $k = 300$) by approximately 200 agents. This means that upon increasing risk avoidance by 40%—by changing the behavior from fully tolerant ($w = 1$) to 60% tolerant ($w = 0.6$)—the system has managed to preclude large outbreaks for a population about 67% larger (from $k = 300$ to $k = 500$).

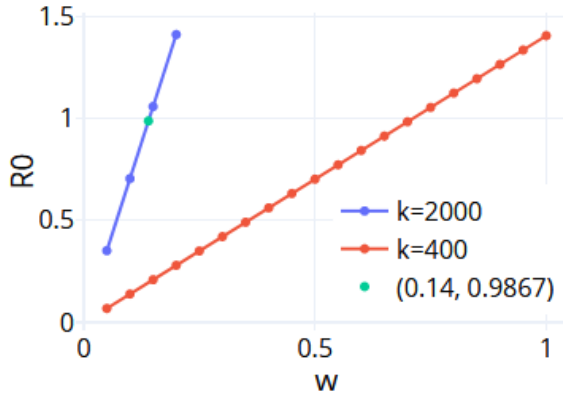
Figure 3.3(b) (x-axis log-scaled) illustrates the impact of the walk rate λ on an epidemic's average duration. Note that 3 out of the 4 curves show scenarios wherein an increase in the walk rate leads to a phase transition (very long duration). In turn, for the curve "$n = 10^5, w = 0.6$" the duration increases slowly even when λ reaches 20. It could be the case of a phase transition still being reached for some even larger λ. However, from (3.19), as $\lambda \to \infty$ the infection rate β converges to $\beta' = (\tau k w)/n$. For the given parameters, it yields a basic reproduction number $R_0' = \beta'/\gamma = 0.0015/0.0019 = 0.79 < 1$. This means that the protective level of $w = 0.6$ will manage to impede long-lasting epidemics *upon any walk rate* λ, in sharp contrast with the scenario where $w = 1$.

3.7.1 Discussion

While the analyses have up to this focused on complete graphs, the proposed model will be expanded in Chapter 4 in order to support more general, degree-uncorrelated networks. Indeed, Figure 3.4 shows results from a simulation indicating how different topologies can decisively change how the epidemic unravels. It compares the

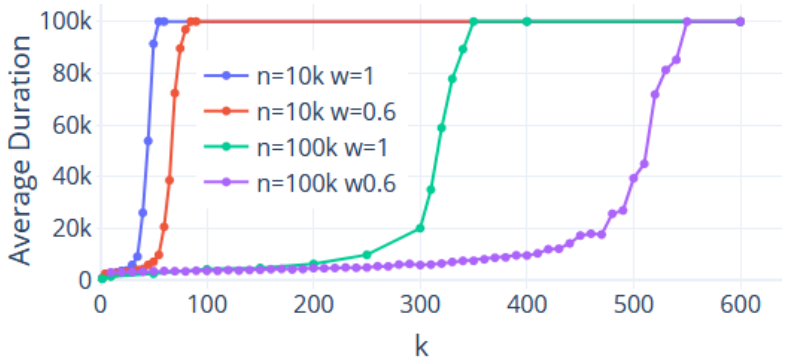


(a) Fraction of infected agents over time ($k = 2000$).

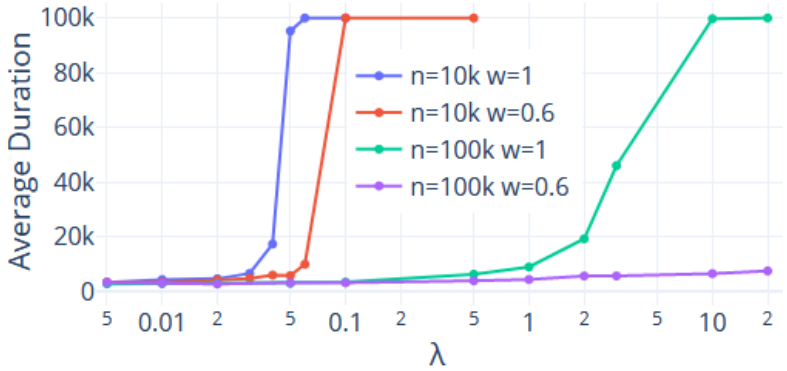


(b) Basic reproduction number as a function of w.

Figure 3.2: Fraction of infected agents as a function of time for different scenarios (a), and the reproduction number R_0 (b) as a function of w when $k = \{400, 2000\}$. For both figures, $\tau = 1; \gamma = 1.9 \cdot 10^{-3}; \lambda = 1;\ n = 10^5$.



(a) Average duration of different epidemics in function of k ($\lambda = 1$).



(b) Average duration of different epidemics in function of λ (x-axis log-scaled; $k = 250$).

Figure 3.3: Average duration of different epidemics in function of the number k of agents (a) and the walk rate λ (b). For both figures, $\tau = 1; \gamma = 1.9 \cdot 10^{-3}$.

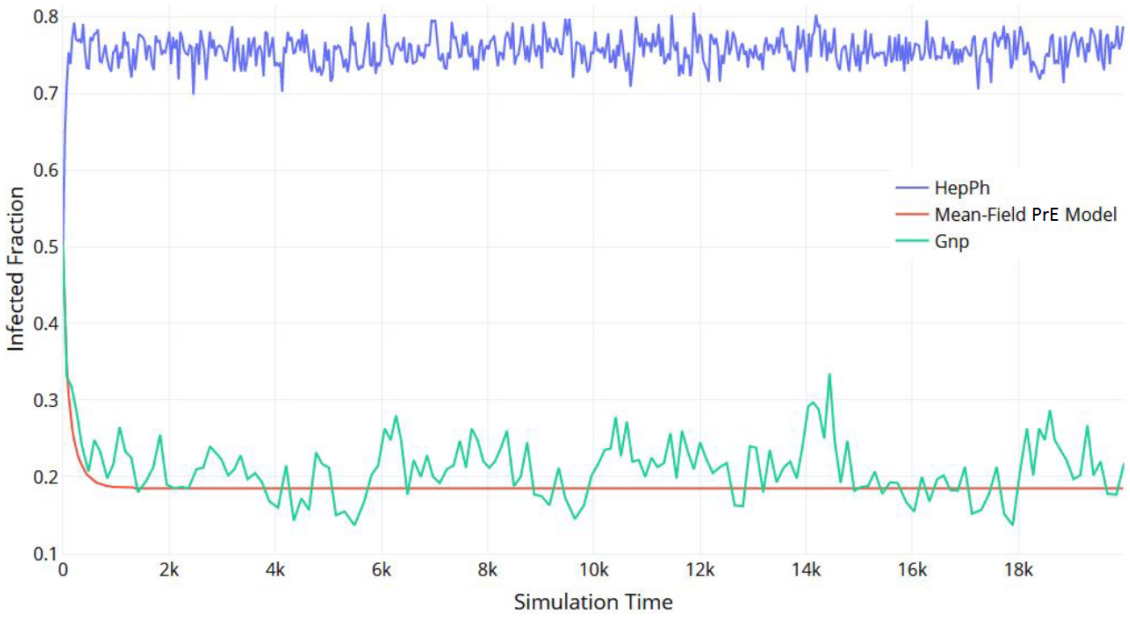


Figure 3.4: Our ODE model (*Mean-Field PrE Model*) fails at predicting epidemics on real networks, while it is relatively accurate for a $G(n, p)$ network with same size and same average degree.

outcome of two equally-parameterized epidemics, diverging from one another due to the network structure. The blue curve denotes a single-run epidemic simulation performed over a real network: the *High-Energy Physics* (Hep-Ph) network [66]. The green curve is an equivalent simulation carried out over a random $G(n, p)$ network *with same size and same average degree* of Hep-Ph. The red curve is the ODE model prediction for this epidemic. Note that, on the one hand, the model is unable to predict the outcome for the real network; on the other, it interestingly delivers a reasonable prospect for $G(n, p)$. In the next chapter we broaden the present scope with the inclusion of non-regular graphs, thus embracing many classes of networks.

3.8 Time-varying protection

Our mobility scheme so far has assumed a fixed protection level: agents repel hostile localities agreeing to some previously established risk tolerance w. One could argue, however, that a *proportional* defensive scheme would likewise express a natural behavior against an outbreak, aligned with the hypothesis that people get more or less responsive to threats in proportion to their own risk perception. While such a perception could result from a combination of many epidemic-related aspects (such as media coverage and number of infected relatives), it is reasonable to link risk awareness to the actual size of the epidemic (i.e., the total number of infectives). That is, we may consider the case where individuals assess an epidemic as more or less dangerous depending on the current number of infected individuals. We hence-

forward consider a *proportionally reactive* approach. To this end, we promote a change in the previous model, described as follows.

3.9 Reactive Model

A straightforward way to obtain a more general responsive setup is to replace w by some function $w(i)$, which in turn will determine different protective levels depending on the current fraction i of I-agents.

We therefore consider that the risk-tolerance from agents is now governed by a function $w(\cdot), 0 \leqslant w(\cdot) \leqslant 1$, such that

$$w(i) = (1 - i)^r, \tag{3.21}$$

for some *rejection force* $r \geqslant 0$. Note that r controls how intensely agents repel SI-contacts as the infectives total increases.

3.9.1 Analysis

We are mainly interested in answering whether there exists a regime for $w(\cdot)$ that ensures short-lived epidemics irrespective of the infection force β and, if so, identifying such a regime. To this end, however, the basic reproduction number R_0 is no longer representative of a proper threshold, as it was in Section 3.6.3. Indeed, because the infection force β now depends on a varying protective level $w(i)$, its value at the endemic regime may also differ from that applied to i_0 (the initial fraction of infectives). More precisely, in Section 3.6.3—wherein the risk-tolerance w were still a fixed parameter—we have defined a constant C (3.17), in order to rewrite the infection force 3.14 as $\beta = 2Cw$ and then notate $R_0 = 2Cw/\gamma$. Here, however, we cannot follow these same lines, i.e. we cannot simply replace w by a function $w(i_0)$, for the aforementioned reason.

We thus resort to another approach in order to understand the protective-behavior impacts on the endemic regime. The first step is to replace w from (3.16) by $w(i)$. Therefore,

$$\frac{di}{dt} = \frac{2\tau\lambda k w(i)}{(2\lambda + \tau)n} si - \gamma i = 2Cw(i)si - \gamma i\,, \tag{3.22}$$

and the steady-state fraction i^* of I-agents is then determined by equaling the above derivative to zero, i.e. i^* must satisfy

$$2Cw(i^*)si^* - \gamma i^* = 0\,. \tag{3.23}$$

To ease notation, we divide all terms by γ and also define another constant, C', such that $C' = \gamma/2C$. Then, (3.23) becomes

$$\frac{w(i^*)si^*}{C'} - i^* = 0\,. \tag{3.24}$$

Now, substituting $w(i^*)$ by the function $w(i^*) = (1 - i^*)^r$ (3.21), and noting that $s = 1 - i^*$, yields

$$\frac{(1 - i^*)^r(1 - i^*)i^*}{C'} - i^* = 0\,, \tag{3.25}$$

thus

$$(1 - i^*)^{r+1} = C'\,, \tag{3.26}$$

which implies that

$$1 - i^* = C'^{\frac{1}{r+1}}\,, \tag{3.27}$$

and hence

$$i^* = 1 - C'^{\frac{1}{r+1}}. \tag{3.28}$$

Since we want to know the regime under which the epidemic vanishes, we must determine when condition $1 - C'^{1/(r+1)} \leqslant 0$ holds. Indeed,

$$1 - C'^{\frac{1}{r+1}} \leqslant 0 \iff 1 \leqslant C'^{\frac{1}{r+1}} \iff 1 \leqslant C'\,. \tag{3.29}$$

(3.29) reveals a strikingly different perspective when compared with that of constant risk-tolerance: the epidemic vanishes only if $C' \geqslant 1$. However, this is exactly the same condition for the epidemic to vanish when no protection effect is present. Remarkably, proportionally reactive protection cannot contribute to the termination of an epidemic, contrary to the constant-protection case (3.18). Note, however, that the proportional protection influences the endemic level, as given by (3.28): larger values for the SI-contact rejection r will lead to smaller endemic levels. Such a conclusion is aligned with previous works that have analysed proportional protection over fully-mixed, unstructured environments [10]: the gradually reactive behavior imposes a saturation level to the epidemic rather than extinguishing it.

Chapter 4

Agent-based Epidemics on Degree-uncorrelated Networks

4.1 Problem Formulation without PrE

We shall now describe the coupling of SIS epidemics [47] with a mobile-agent environment, as considered in this work. At this point, we are not yet considering protective mobility (which will be included later in the text), but only uniform random walks instead. Table 4.1 lists the main terms and parameters to be used throughout this section.

Mobility and contact pattern. Consider an undirected network $G = (V, E)$ with node and edge set given by V and E, respectively, where $n = |V|$ denotes its size. Also, consider a set A of $K = |A|$ agents. Here, again, A is assumed to be *sparse* i.e. $K \ll n$. An exact definition for sparsity as considered herein is later provided in Section 4.3.4. Agents move across G according to uniform, continuous-time random walks, hereafter named *Simple Random Walks* (SRW), as follows. First, we associate a self-loop with each node so that an agent about to walk is always allowed to take no action—by choosing its current location—or to effectively go somewhere else. Note that this characterizes a *lazy* random walk. More specifically, let $v_j(t) \in V, j \in A$ denote the location of agent j at time $t \in \mathbb{R}_+$, and let $N(v) = \{u \in V \mid (u, v) \in E\}$ denote the *closed* neighborhood of node v. Also, assume the starting location of j, $v_j(0)$, is uniformly chosen at random from $V, \forall j \in A$. Thereafter, the residence time of j at any given node is exponentially distributed with rate $\lambda > 0$, the walk rate. Upon movement at time t, j's next-hop is uniformly chosen at random from $N(v_j(t))$. Hops are assumed to occur instantaneously.

Any two agents j and l are said to be in contact with one another *iff* both of them are located at the same node, i.e. if $v_j(t) = v_l(t)$. Note that, on top of a static topology (imposed by G), agents produce another, time-varying *network of contacts*

Table 4.1: Symbols & Terminologies

S-/I-agent	Susceptible / Infected agent, respectively.
$G = (V, E)$	Undirected network wherein mobile-agents walk.
n	Network size; $n = \|V\|$.
$N(v)$	Set of v's neighbors *including* v, $v \in V$.
B	Set of degree blocks from G; $B = \{b_1, b_2, ..., b_{\|B\|}\}$.
$d(v)$	v's degree, $v \in V$; $d(v) = \|N(v)\| = b$, for $b \in B$.
V_b	Set of degree-b nodes; $V_b = \{v \in V \mid d(v) = b\}$.
SI-contact	Pairwise contact between S- and I-agents at v.
A, K	Set and number of agents, respectively; $K = \|A\|$.
$S(t), I(t)$	Set of S-/I-agents at *continuous* time t.
S, I	Shorthand for $\|S(t)\|$ and $\|I(t)\|$; $S + I = K$
s, i	S/K and I/K, respectively, hence $s + i = 1$.
I_0, i_0	Number and fraction of I-agents at $t = 0$, respectively.
S_v, I_v	Number of S-/I-agents currently at node v, respectively.
S_b, I_b	Number of S-/I-agents currently at block b, respectively.
τ	Pathogen's transmissibility rate.
λ, γ	Walk rate, and recovery rate, respectively.
λ_b	Exit rate from block b; $\lambda_b = ((b-1)/b)\lambda$.
v_b	A node with degree b or a node from block b.
p_b	Fraction of nodes with degree b (i.e., from block b).
n_b	Number of degree-b nodes; $n_b = np_b$; $\sum_b n_b = n$.
q_b	Prob. rand. chosen link points to block b; $q_b = bp_b/\langle b \rangle$.
$\langle b \rangle$	Network's average degree; $\langle b \rangle = \sum_b bp_b$.
$\langle b^2 \rangle$	2^{nd} moment of the degree distrib.; $\langle b^2 \rangle = \sum_b b^2 p_b$.
$\langle k \rangle_b$	Expected density of agents in block b; $\langle k \rangle_b = bp_b/\langle b \rangle$.
$\langle \lambda_b \rangle$	Expected exit rate; $\langle \lambda_b \rangle = \sum_b \lambda_b \langle k \rangle_b = \lambda(1 - 1/\langle b \rangle)$
$\langle K \rangle_b$	Expected population in block b; $\langle K \rangle_b = Kbp_b/\langle b \rangle$.
$\langle b \rangle_A$	Agents' expected block; $\langle b \rangle_A = \sum_b b\langle k \rangle_b = \langle b^2 \rangle/\langle b \rangle$.
$\langle a \rangle$	Expected agglomeration; $\langle a \rangle = K\langle b^2 \rangle/(n\langle b \rangle^2)$.

characterized by a collection of isolated cliques.

Epidemic state and infection. Each agent is in one of two states, namely "S" (susceptible) or "I" (infected). Let $c_j(t) \in$ { "S","I" } denote the epidemic state of agent j at time $t \geqslant 0$. Also, let $S(t) = \{j \in A | c_j(t) =$"S"$\}$ and $I(t) = \{j \in A | c_j(t) =$"I"$\}$ denote the set of susceptible and infected agents at time t, respectively, such that $S(t) \cup I(t) = A$ and $S(t) \cap I(t) = \varnothing$ for all t. Disease spreads through direct contact between S- and I-agents—hereafter denoted *SI-contact*—, with infection probability proportional not only to the duration of such a contact but also to the number of infectives the S-agent is in contact with. Since an SI-contact may arise when either S- or I-agents enter a *hostile node* i.e. a node containing at least one agent from a different epidemic state, both cases must be described separately, as follows.

- *S-agent j enters a node v hosting I_v I-agents.* In this case, each infective throws an exponential coin Y with parameter $\tau > 0$, the pathogen's *transmissibility rate*, independently. Therefore, a total of $\{Y_1, Y_2, ..., Y_{I_v}\}$ coins are generated in function of j. Then, j gets infected *iff* any Y_i fires before it leaves the node *and* before I-agent i leaves as well.

- *I-agent j enters a node v hosting S_v S-agents.* Here the incoming infective throws S_v exponential coins $\{Y_1, Y_2, ..., Y_{S_v}\}$ with parameter $\tau > 0$, one for each S-agent at v. Infection then happens as in the previous item.

Recovery. Once infected, an individual so remains for a certain time window, recovering right after. The elapsed time until an I-agent recovers is exponentially distributed with rate $\gamma > 0$, the *recovery rate*. Recoveries are independent from agent to agent.

Note that any S-agent, once infected, initiates an infection event for each S-agent still remaining at v. Likewise, when an I-agent recovers, it is exposed to new transmission events produced by each I-agent remaining at v.

4.2 Problem Statement III

Our third goal is to establish a fluid model that predicts the average behavior of epidemics governed by the aforementioned laws, where degree-uncorrelated networks are considered but no risk-awareness. Later in Chapter 5 we further extend the problem at hand by finally encoding protective behavior with degree-uncorrelated networks.

4.3 Key Definitions

In order to properly describe our preliminary predictor, let us first revisit and also introduce some terminology, as follows.

4.3.1 Degree-block approximation

Often, a network $G = (V, E)$ may be partitioned into degree-based subsets so that many nodes are treated altogether based on their degree, a procedure called *degree-block approximation* [1], which we have already defined in Section 2.2.1. We now formalize such blocks as follows. Let $B = \{b_1, b_2, ...b_{|B|}\}, b_1 < b_2 < ... < b_{|B|}$, denote the set of all degree blocks occurring in G. Each $b \in B$ informs that at least one node with degree b occurs in G. For instance, if $b_1 = 3$ then $|\{v \in V \mid d(v) = 3\}| \geqslant 1$.

4.3.2 Block Probability

In a network that lacks degree correlations, the probability that a randomly chosen link points from a node with degree b' to one with degree b is independent of b' [48]. Consequently, if n_b and p_b are the number and the fraction of nodes with degree b, respectively (hence $n_b = np_b$), $\langle b \rangle$ is the network's average degree and $m = |E|$ is the total number of undirected edges (links), the probability q_b that a randomly chosen link points to *any* degree-b node is

$$q_b = n_b \frac{b}{2m - n} = np_b \frac{b}{2m - n} = \frac{bp_b}{\langle b \rangle}. \tag{4.1}$$

Note that self-loops (made present for all nodes) must be counted only once, hence the denominator is $2m - n$ instead of simply $2m$.

4.3.3 b-Expected Density/Population

Given a node v and an agent j, it is a well-known result that, for a stationary distribution π of a SRW, the stationary probability π_v that j is at v corresponds to v's degree share, i.e. $\pi_v = \frac{d(v)}{2m-n}$ (self-loops considered). Consequently, the stationary probability $\langle k \rangle_b$ that j is at *any* degree-b node (i.e. any node v such that $d(v) = b$), or the *b-expected density*, is

$$\langle k \rangle_b = \sum_{v \in V_b} \pi_v = \frac{b}{2m - n} n_b = \frac{bnp_b}{2m - n} = \frac{bp_b}{\langle b \rangle}. \tag{4.2}$$

(4.2) expresses a not an obvious—and not always valid—relation: the expected density of agents within some block matches the probability that a randomly chosen link points to that same block (4.1) i.e. $\langle k \rangle_b = q_b$. *This holds due to the lack of degree correlations.* If not for this, q_b's definition would be different (as it depends on how degrees correlate), whereas $\langle k \rangle_b$'s would remain as is.

The expected population $\langle K \rangle_b$ of agents within a specific block b, or the *b-expected population*, follows immediately from (4.2), and reads

$$\langle K \rangle_b = K \frac{bp_b}{\langle b \rangle}. \tag{4.3}$$

4.3.4 Sparsity

We now establish an explicit sparsity threshold. We say a scenario is sparse if

$$\max_b (\langle K \rangle_b / n_b) \leqslant 0.1. \tag{4.4}$$

The above threshold means, on the one hand, that every next-hop is likely to be vacant; on the other, it also stands at a safe margin from random fluctuations that could sensibly violate the premise on pairwise encounters only.

4.3.5 Agents' Expected Block

A node freely chosen uniformly at random *has* expected degree $\langle b \rangle = \sum_{v \in V} d(v)\frac{1}{n}$, but a *next-hop* (also a node) randomly chosen by a set A of agents *belongs to* an expected block $\langle b \rangle_A$, which reads

$$\langle b \rangle_A \triangleq \mathbb{E}_\pi[d(\cdot)] = \sum_{v \in V} d(v)\pi_v. \tag{4.5}$$

Despite being a well-known result, Eq. 4.5 is hardly tractable in practice as it lacks a closed form. By explicitly accounting for the different degree blocks, our next result not only resolves Eq. 4.5 but even raises perspectives beyond the scope of this paper. Briefly, we prove that an SRW and the degree function $d(\cdot)$ may be conveniently combined in order to converge in expected value to *any desired moment of a network's degree distribution*, as follows.

Proposition 1 (Law of convergence to the n-th moment) *Let $d : V \to \mathbb{N}$ denote the degree function, to be measured up from each node via SRW in an undirected network $G = (V, E)$. If such a walk has a unique stationary probability $\pi_v > 0$ of being at any node $v \in V$, then $\mathbb{E}_\pi\left[d(\cdot)^{n-1}\right] / \mathbb{E}_\pi\left[d(\cdot)^{-1}\right] = \langle b^n \rangle$, $\forall n \in \mathbb{Z}_+$.*

Proof:

$$\mathbb{E}_\pi\left[d(\cdot)^{n-1}\right] \triangleq \sum_{v \in V} d(v)^{n-1}\pi_v = \sum_b \sum_{v \in V_b} b^{n-1}\pi_v \overset{Eq.\ 4.2}{=}$$

$$\sum_b b^{n-1}\langle k \rangle_b = \sum_b b^{n-1}\frac{bp_b}{\langle b \rangle} = \frac{1}{\langle b \rangle}\sum_b b^n p_b = \frac{\langle b^n \rangle}{\langle b \rangle}.$$

Consequently, if $n = 0$ then

$$\mathbb{E}_\pi\left[d(\cdot)^{n-1}\right] = \mathbb{E}_\pi\left[d(\cdot)^{-1}\right] = \frac{1}{\langle b \rangle}, \text{ hence}$$

$$\frac{\mathbb{E}_\pi\left[d(\cdot)^{n-1}\right]}{\mathbb{E}_\pi\left[d(\cdot)^{-1}\right]} = \frac{\langle b^n \rangle}{\langle b \rangle}\langle b \rangle = \langle b^n \rangle, \text{ and the claim holds.} \quad \square$$

Remarkably, Proposition 1 provides a closed form for Eq. 4.5, which now reads

$$\langle b \rangle_A \triangleq \mathbb{E}_\pi\left[d(\cdot)\right] = \frac{\langle b^2 \rangle}{\langle b \rangle}. \tag{4.6}$$

To the best of our knowledge, the above derivation, which leads to a general closed form for $\mathbb{E}_\pi\left[d(\cdot)\right]$—and is valid for any connected, undirected network—is new. Indeed, the above result has been demonstrated only for the *particular case* of

the configuration model [49], which breeds *only degree uncorrelated networks* (or negligibly correlated). Conversely, through Proposition 1 we prove that such a ratio applies to more general networks, with *arbitrary degree correlations* (e.g. a star network).

Eq. 4.6 exposes a universal property of uniform random walks, so far only observed empirically: the average degree $\langle b \rangle$ for heterogeneous networks tends to be largely overestimated when learned via SRW [67]. Proposition 1 not only proves that this is *always* the case, but also states the exact value being approached *irrespective of degree correlations*. By noticing that

$$\frac{\langle b^2 \rangle}{\langle b \rangle} = \frac{\langle b^2 \rangle - \langle b \rangle^2 + \langle b \rangle^2}{\langle b \rangle} = \langle b \rangle + \frac{\mathrm{Var}(b)}{\langle b \rangle}$$

the extra term—directly proportional to the network's heterogeneity—becomes explicit.

Noteworthy, $\langle b^2 \rangle / \langle b \rangle$ may be considered a special quantity in the study of networks, with several other physical interpretations and/or applications apart from mobile-agent contexts. These include the network's heterogeneity coefficient [68] and the Molloy-Reed criterion for the existence of a giant connected component [69]. For the particular case of degree-uncorrelated networks, it also accounts for the largest eigenvalue of the adjacency matrix of power-law networks with exponent $2 < \gamma < 5/2$ [70] and the average degree of a node's nearest neighbors [49, 68].

Although beyond the present scope, it should be briefly mentioned that Proposition 1 also accounts for a promising reference on the design of estimators for the network moments via random walks.

4.3.6 Expected Agglomeration

Another key quantity for later discussion is the expected agglomeration $\langle a \rangle$ of agents per node, which we define as

$$\langle a \rangle = \sum_b \frac{\langle K \rangle_b}{n_b} \langle k \rangle_b = \frac{K \langle b^2 \rangle}{n \langle b \rangle^2} = \frac{K \langle b \rangle_A}{n \langle b \rangle}. \tag{4.7}$$

4.3.7 Exit Rate

Self-loops capture the ability every agent has of staying at its current node whenever its walk event triggers. This begs a clear distinction between the walk rate and the *exit rate* i.e., the rate at which an agent's decision on its next hop effectively takes it elsewhere. Consequently, every node with degree b induces its own exit rate λ_b, defined as

$$\lambda_b = \frac{b-1}{b} \lambda. \tag{4.8}$$

4.3.8 Expected Exit Rate

Though a single walk rate λ applies to the entire network, the same is not true with respect to the exit rate, which is block dependent (4.8). It will later prove relevant to also consider the *expected exit rate* $\langle \lambda_b \rangle$, which reads

$$\langle \lambda_b \rangle = \sum_b \lambda_b \langle k \rangle_b = \lambda \left(1 - \frac{1}{\langle b \rangle} \right). \tag{4.9}$$

(4.9) sheds light on how the inclusion of self-loops can substantially improve the epidemic prediction accuracy. Indeed, the very definition of q_b (4.1) implicitly assumes that, from anywhere in the network, an agent is allowed to go to any other node, *including its current node*. Such is only the case in the presence of self-loops. Otherwise, an *agent's exit rate would be poorly estimated, should the average degree* $\langle b \rangle$ *be small* (4.9).

4.3.9 Infection Probability

Encounters in a sparse regime occur pairwise only, w.h.p. Since exponentials are *memoryless*, two agents in SI-contact at some node v with degree b may be assumed to have arrived at the same time. Disease thus spreads with some probability σ_b that the I-agent's transmission event fires (at rate τ) prior to some exit event (at rate $2\lambda_b$, since both agents may exit v). We show in Appendix A that

$$\sigma_b = \frac{\tau}{2\lambda_b + \tau}. \tag{4.10}$$

4.3.10 Expected Infection Probability

It will also be important to discuss the infection probability in terms of its expected value across the different degree blocks, as follows:

$$\langle \sigma_b \rangle = \sum_b \sigma_b \langle k \rangle_b. \tag{4.11}$$

4.4 General Block-Level System

Our general-purpose predictor is

$$\begin{cases} \dfrac{dI_b}{dt} = (I - I_b)\lambda q_b - I_b \lambda \overline{q_b} + I_b \lambda_b \dfrac{S_b}{n_b}\sigma_b + S_b \lambda_b \dfrac{I_b}{n_b}\sigma_b - \gamma I_b, \\ \dfrac{dS_b}{dt} = (S - S_b)\lambda q_b - S_b \lambda \overline{q_b} - I_b \lambda_b \dfrac{S_b}{n_b}\sigma_b - S_b \lambda_b \dfrac{I_b}{n_b}\sigma_b + \gamma I_b. \end{cases} \tag{4.12}$$

(4.12) states that the number of S-/I-agents within some block b varies with time according to five main quantities. Let us walk through these from the I-agent's perspective (the same ideas apply for S-agents):

- $(I - I_b)\lambda q_b$: the number of I-agents from outside b that, walking at rate λ, hop into b with probability q_b;
- $I_b\lambda\overline{q_b}$: the number of I-agents within b that, walking at rate λ, leave that block with probability $\overline{q_b} = 1 - q_b$;
- $I_b\lambda_b\dfrac{S_b}{n_b}\sigma_b$ and $S_b\lambda_b\dfrac{I_b}{n_b}\sigma_b$: the number of I-(resp. S-) agents already within b that, walking at an effective rate λ_b, (i) encounter a total of S_b/n_b (resp. I_b/n_b) S-(resp. I-) agents and (ii) promote a new infection (resp. get infected) with probability σ_b;
- γI_b: the number of I-agents within b that recover at rate γ.

With straightforward calculations, (4.12) reduces to

$$\begin{cases} \dfrac{dI_b}{dt} = \lambda(Iq_b - I_b) + \dfrac{2\lambda_b\sigma_b}{n_b}S_bI_b - \gamma I_b \\ \\ \dfrac{dS_b}{dt} = \lambda(Sq_b - S_b) - \dfrac{2\lambda_b\sigma_b}{n_b}S_bI_b + \gamma I_b, \end{cases} \tag{4.13}$$

which in terms of densities i and s reads

$$\begin{cases} \dfrac{di_b}{dt} = \lambda(iq_b - i_b) + \dfrac{2\lambda_b\sigma_bK}{n_b}s_bi_b - \gamma i_b \\ \\ \dfrac{ds_b}{dt} = \lambda(sq_b - s_b) - \dfrac{2\lambda_b\sigma_bK}{n_b}s_bi_b + \gamma i_b. \end{cases} \tag{4.14}$$

The evolution of infectives is thus obtained through summing up (4.13) for each block i.e.,

$$\frac{dI}{dt} = \frac{2}{n}\sum_b \frac{\lambda_b\sigma_bS_bI_b}{p_b} - \gamma I, \tag{4.15}$$

or, in density terms,

$$\frac{di}{dt} = \frac{2K}{n}\sum_b \frac{\lambda_b\sigma_bs_bi_b}{p_b} - \gamma i. \tag{4.16}$$

Note, however, that our predictor is still required to solve (4.13) since S_b and I_b must be determined for each block. The relevance of (4.15) and (4.16) is made more clear in the next section, where we show the existence of a special condition under which blocks may be treated independently from one another, leading to a huge analytical simplification and consequent computational speedup on the numerical solution.

4.5 Matching Density Condition (MDC)

This is perhaps our most surprising result. Depending on the input parameters, dynamics over networks with *arbitrary degree distribution* may be accurately predicted not by a system but by *one single equation*. We shall now describe it from the S-agents standpoint (the same ideas apply for I-agents).

4.5.1 State-level mixing & MDC

The key observation is: if every S-agent j performs, with high probability (w.h.p), sufficiently many hops prior to get infected, *to the point of approaching the network's mixing time*, then the probability that j is at any node v *still as an S-agent* approaches π_v. We call this phenomenon *state-level mixing*, when either S- or I-agents—not necessarily both—do approach the mixing time w.h.p prior to changing their epidemic state. The immediate consequence, on the one hand, is that $\mathbb{E}[S_b/S] \equiv \langle k \rangle_b \ \forall b$. On the other, since every K_b is partitioned as $S_b + I_b$, then $\mathbb{E}[I_b/I] \equiv \langle k \rangle_b$ would *forcefully* hold as well (irrespective of whether I-agents do mix at state level), thus the whole dynamics exhibit the *Matching Density Condition* i.e.

$$\mathbb{E}[S_b/S] \equiv \mathbb{E}\left[I_b/I\right] \equiv \langle k \rangle_b \ \ \forall b \in B. \tag{4.17}$$

(4.17) states that, upon MDC, not only are the *relative* densities of susceptible/infected agents within each block state invariant, but also that these match the stationary density of agents at b.

MDC allows us to express every S_b and I_b strictly in terms of the variables S and I and the block-level constants $\langle k \rangle_b$. That is, provided MDC, then $S_b = \langle k \rangle_b S$ and $I_b = \langle k \rangle_b I \ \forall b \in B$. In what follows we describe (i) how to identify MDC beforehand, and (ii) the epidemic dynamics with MDC and its many striking consequences.

4.5.2 Identifying MDC

To determine in advance whether S-agents do predominantly mix is a hard task, as their states also depend on their interactions with I-agents. On the flip side, it is quite straightforward to achieve this sense from an I-agent perspective. Indeed, once infected, the expected number $\langle h_I \rangle$ of hops an I-agent performs until recovering is a simple relation between its expected exit rate $\langle \lambda_b \rangle$ and the recovery rate γ. It can be shown that

$$\langle h_I \rangle = \langle \lambda_b \rangle / \gamma. \tag{4.18}$$

When the bound on the mixing time for an input network is verifiable [71], a trivial check on whether $\langle h_I \rangle$ stands within said bound suffices. Briefly, let π denote the

stationary distribution of a single random walk in a network with spectral gap δ. Now, for $\pi_0 = \min \pi_i \in \pi$, I-agents are, on average, at most at an $\varepsilon > 0$ distance from π if

$$\langle h_I \rangle \geqslant \frac{\log(1/(\pi_0 \varepsilon))}{\delta}. \tag{4.19}$$

At this point it is clear that one may not always identify in advance whether MDC applies. Indeed, its computation not only requires previous knowledge on the network's degree distribution but also may be costly or even time-prohibitive.

4.5.3 Dynamics with MDC

Consider an epidemic process that unfolds upon MDC. Therefore, $S_b = \langle k \rangle_b S$ and $I_b = \langle k \rangle_b I$. Substituting in Eq. 4.15 yields

$$\frac{dI}{dt} = \frac{2\langle \lambda_b \rangle \langle \sigma_b \rangle \langle b^2 \rangle}{n \langle b \rangle^2} SI - \gamma I, \tag{4.20}$$

which in terms of densities i and s becomes

$$\frac{di}{dt} = \frac{2\langle \lambda_b \rangle \langle \sigma_b \rangle K \langle b^2 \rangle}{n \langle b \rangle^2} si - \gamma i \overset{\text{Eq. 4.7}}{=} 2\langle \lambda_b \rangle \langle \sigma_b \rangle \langle a \rangle si - \gamma i. \tag{4.21}$$

(4.20) and (4.21) are in many ways a striking result. They impact several aspects of the spreading dynamics at hand, which we elaborate as follows.

• *Block Independence.* Block-level variables vanish upon MDC. This represents a huge analytical and numerical simplification: whereas the size complexity for the number of equations of a block-level solution is $\Theta(|B|)$, MDC is $\Theta(1)$.

• *Degree Distribution.* Interestingly, not only S_b and I_b disappear, but also q_b no longer plays a role as well, due to inter-block movements being neglected (even though these still occur in practice), thus the premise of no degree correlation drops. Moreover, because at least one epidemic partition (either S- or I-agents) reach state-level mixing, their density at each block is equally distributed across the nodes from that block w.h.p. In this case, (4.20) and (4.21) do apply *irrespective of the network's degree distribution.*

• *Sparsity.* Dynamics under MDC only depends on how the rates for walk, transmissibility and recovery relate to one another, thus the sparsity premise drops. In this work, a dependence on the number of agents still exists simply because the infection probability σ_b within each block, as we defined (4.10), assumes pairwise encounters only. A dense regime would require another definition for σ_b to be conceived—a task beyond our present scope—and then plugged back directly into (4.20) and (4.21).

• *Structural Influence.* General key aspects of the underlying network are now

made explicit. Indeed, even those scenarios where MDC does not hold are in fact subject to the same network structure and, consequently, to its influence. Beyond previously evidenced impact of its size (4.15), another major topological aspect have now became pronounced: larger degree heterogeneity ($\langle b^2 \rangle / \langle b \rangle^2$) boosts the spreading. It means that regularly well-connected topologies are best for disease containment i.e., safer urban scenarios should ideally provide direct path between any two regions but at equivalent proportions, in a fashion similar to regular graphs.

Of particular interest is how this feature compares to static contact models, where nodes encode individuals rather than sites [48]. Therein, larger $\langle b \rangle$ propels the outbreak, which is not the case in our model. Remarkably though, the impact of degree heterogeneity in both models turns out to be the same: to intensify the spreading.

Last, note that the expected agglomeration (4.7) has naturally arose into (4.21) in replacement to an otherwise block-level system, in agreement with intuition.

• *Classical SIS.* It is now also evident that the classical, homogeneous-mixing SIS model (2.7) does actually correspond to *a particular case of a particular case* of the herein proposed model. Indeed, MDC itself is a particular scenario which embeds a *force of infection* β defined as

$$\beta = 2\langle \lambda_b \rangle \langle \sigma_b \rangle \langle a \rangle, \tag{4.22}$$

so that (4.21) becomes $di/dt = \beta si - \gamma i$, which in turn is further reduced to the homogeneous-mixing SIS (2.7) by simply making $\beta = \tau$.

Chapter 5

Protection Effects on Degree-uncorrelated Networks

5.1 Protection Model

We now include the protective behavior artifact into the model presented in the previous chapter. *Protection here is modeled in the exact same way described in Section 3.4.* With that protective model at hand, we proceed to establish our ultimate problem statement.

5.2 Problem Statement IV

Our ultimate goal is to generalize the fluid model from Chapter 4 in order to predict the progress of epidemics where agents are now biased as described in this section.

5.3 Theoretical analysis

The PrE model proposed in Section 3.4 is now coupled with the epidemic model presented in Section 4.1 in order to evaluate key-aspects, namely (i) the SI-contact rate (a fundamental model parameter), (ii) prediction on the dynamics, and (iii) the basic reproduction number R_0. As in the case of complete graphs, w_s and w_i are block-independent, we will simply consider a single, equivalent protective level w given in terms of their mean, as described in (3.6).

5.4 SI-contact rate

The SI-contact rate within each block now depends on the PrE parameter w (3.6). To make further progress on how exactly w influences such a rate, we must first describe the "two-stage migration" approach to compute the probabilities of SI-contacts, as follows.

5.4.1 Two-stage Migration

From a strictly probabilistic standpoint, the entry of an agent j in a node v_b from block b may be thought of as a sequence of *two chained events*: (i) j chooses a block b with probability q_b and then (ii) once b is established, j somehow chooses some v_b. Let us denote α_{S_b} and α_{I_b} the SI-contact rates within b produced, respectively, by the walks of those S-/I-agents already within b that step into hostile nodes also from b. Therefore, $\alpha_{S_b} = S_b\lambda_b\rho_{S_b}$ and $\alpha_{I_b} = I_b\lambda_b\rho_{I_b}$, where ρ_{S_b} and ρ_{I_b} are, respectively, the time-varying probabilities that an S-/I-agent from b provokes an SI-contact within b. We thus need to find the instantaneous values for ρ_{S_b} and ρ_{I_b}. When random walks are uniform (i.e. $w = 1$), then simply $\rho_{S_b} = I_b/n_b$ and $\rho_{I_b} = S_b/n_b$, so that $\alpha_{I_b}\sigma_b$ and $\alpha_{S_b}\sigma_b$ yield the 3rd and 4th terms of Equation 4.12, respectively.

In the presence of PrE however, the two-stage separation comes in handy: we may still consider blocks as being chosen by j as described in (i) since, in the general case, sparsity implies that the probability density deducted from hostile nodes and equally redistributed across safe sites does not meaningfully change the likelihood of each block in comparison to the others. However, once a block b is elected, j's final residence can no longer be drawn uniformly out of n_b possibilities since j is now prone to safe localities. In this case, an appropriate stage-(ii) behavior is described as follows.

5.4.2 SI-contact Rate with PrE

Firstly, note that those I_b I-agents within b occupy I_b out of n_b nodes w.h.p., due to sparsity. In turn, each such location is avoided by S-agents with some bias w, whereas the other $n_b - I_b$ nodes from b are chosen with unitary weight. Thus, the instantaneous probability that the S-agent j provokes an SI-contact at block b is simply $\rho_{S_b} = I_b w/(I_b w + (n_b - I_b)) = I_b w/(I_b(w-1) + n_b)$, but since $I_b(w-1) \ll n_b$ (due to sparsity), then $\rho_{S_b} \approx I_b w/n_b$ so that

$$\alpha_{S_b} = S_b \lambda_b \frac{I_b w}{n_b}. \tag{5.1}$$

Analogously, $\alpha_{I_b} = I_b \lambda_b \dfrac{S_b w}{n_b}$. Note that $\alpha_{S_b} \equiv \alpha_{I_b}$. These are fundamental parameters for the model, and also the main modification required in relation to the previous section.

5.5 Evolution of infectives

By applying (5.1) in (4.13) with already shown simplifications, the dynamics with PrE are captured by

$$\begin{cases} \dfrac{dI_b}{dt} = \lambda(Iq_b - I_b) + \dfrac{2\lambda_b \sigma_b w}{n_b} S_b I_b - \gamma I_b \\ \\ \dfrac{dS_b}{dt} = \lambda(Sq_b - S_b) - \dfrac{2\lambda_b \sigma_b w}{n_b} S_b I_b + \gamma I_b, \end{cases} \tag{5.2}$$

which in terms of densities i and s reads

$$\begin{cases} \dfrac{di_b}{dt} = \lambda(iq_b - i_b) + \dfrac{2\lambda_b \sigma_b K w}{n_b} s_b i_b - \gamma i_b \\ \\ \dfrac{ds_b}{dt} = \lambda(sq_b - s_b) - \dfrac{2\lambda_b \sigma_b K w}{n_b} s_b i_b + \gamma i_b. \end{cases} \tag{5.3}$$

When MDC holds, then

$$\frac{dI}{dt} = \frac{2\langle\lambda_b\rangle\langle\sigma_b\rangle\langle b^2\rangle w}{n\langle b\rangle^2} SI - \gamma I, \tag{5.4}$$

which in terms of density reads

$$\frac{di}{dt} = \frac{2\langle\lambda_b\rangle\langle\sigma_b\rangle K\langle b^2\rangle w}{n\langle b\rangle^2} si - \gamma i, \tag{5.5}$$

so that the force of infection becomes

$$\beta = \frac{2\langle\lambda_b\rangle\langle\sigma_b\rangle K\langle b^2\rangle w}{n\langle b\rangle^2} \overset{\text{Eq. 4.7}}{=} 2\langle\lambda_b\rangle\langle\sigma_b\rangle\langle a\rangle w, \tag{5.6}$$

with the classical SIS still being a particular case of this even more general model, again by simply making $\beta = \tau$.

5.6 Reproduction number

The basic reproduction number $R_0 = \beta/\gamma$ is a classical metric that considers the pathogen's spreading potential, and is used to indicate whether the epidemic will die out shortly ($R_0 < 1$) or long last among population ($R_0 > 1$) [47].

Within the context of PrE, a natural question arises: is there a range of values for w that *ensures* $R_0 < 1$? In order to answer this question, let us first consider a constant $C = 2\langle\lambda_b\rangle\langle\sigma_b\rangle\langle a\rangle$ so that $\beta = Cw$. The problem is then to determine which values for w yield $(Cw)/\gamma < 1$. Trivially, if $C/\gamma < 1$ then the epidemic is led to extinction even in the absence of protective efforts. Conversely, if $C/\gamma > 1$, then

$$R_0 < 1 \iff w < \frac{\gamma}{C} \iff w < \frac{\gamma}{2\langle\lambda_b\rangle\langle\sigma_b\rangle\langle a\rangle}. \tag{5.7}$$

(5.7) provides good news: it states that, for the epidemic problem at hand, there always exists some $w > 0$ that imposes $R_0 < 1$, i.e. the joint engagement of S- and I-agents may forcefully prevent endemic steady-states.

5.6.1 Asymptotic regimes

Values for w that satisfy (5.7) arise as a function of other model parameters, particularly the walk rate λ and transmissibility rate τ (within $\langle\sigma_b\rangle$). This introduces the question: is there a regime for w under which $R_0 < 1$ *even if* λ (resp. τ) *is arbitrarily large*, for a fixed τ (resp. λ)? This regime can be identified by considering the asymptotic behavior of λ and τ, one at a time, on the infection force β. Applying (4.10) on (5.6) yields

$$\lim_{\lambda\to\infty} 2\langle\lambda_b\rangle\langle\sigma_b\rangle\langle a\rangle w = \lim_{\lambda\to\infty} \frac{2\tau\langle\lambda_b\rangle\langle a\rangle w}{2\langle\lambda_b\rangle + \tau} = \tau\langle a\rangle w = \beta', \tag{5.8}$$

which yields a basic reproduction number $R_0 = \beta'/\gamma$. Thus,

$$R_0 < 1 \iff w < \frac{\gamma}{\tau\langle a\rangle}. \tag{5.9}$$

Interestingly, any w satisfying (5.9) will manage to extinguish the epidemic *irrespective of the walk rate* λ. Indeed, larger walk rates increase SI-contacts per unit time, but also reduce their duration as well. In similar fashion, as $\tau \to \infty$ (for a fixed λ), β converges to $\beta' = 2\langle\lambda_b\rangle\langle a\rangle w$. Here, $\beta' < \gamma$ is met case $w < \frac{\gamma}{2\langle\lambda_b\rangle\langle a\rangle}$. Any such w will succeed at imposing short-lived epidemics *irrespective of the transmissibility rate* τ.

5.7 Numerical results

In what follows we present simulation results that validate the theoretical analysis from Section 5.3. A discrete-event simulator was designed and implemented [72] in order to generate performance metrics concerning the impacts of PrE and other model parameters. More details on the simulator and its implementation are provided later in Section 5.8. All results to follow assume $i_0 = 0.5$. The population of early spreaders was chosen uniformly at random from the agent set. The starting position of each agent within the network is established as follows. First, every agent is designated to enter some node uniformly chosen at random. Then, every agent is required to walk for a certain time window, and only after that the set of early infectives is finally established. This procedure eliminates an otherwise strong and unrealistic assumption of having all agents initially stopped and without any walk event scheduled prior to the epidemic to begin.

Figures 5.1 and 5.2 show the fraction of infected agents over time for different levels of PrE in a Barabási-Albert network and in a G(n,p) network, respectively. Each given w yields three different curves in both plots: the numerical simulation from one single run and the prediction on the fraction of infectives, as given by the block-level system (5.3) and the network-level master equation (5.5). Note that both models capture with high accuracy the average dynamics for each protection level, since the epidemic's early stage up to its stationary regime. Moreover, it is clear how w decisively changes the epidemic's outcome. In particular, in Figure 5.1 the protective level $w = 0.05$ has managed to hinder the epidemic, leading to a disease-free steady state. The same occurs in Figure 5.2 when $w = 0.1$.

Fig. 5.3 illustrates the impact of the walk rate λ on an epidemic's average duration. Note that 3 out of the 4 curves show scenarios wherein an increase in the walk rate leads to a phase transition (very long duration). In turn, for the curve "$w0.65$" the duration increases slowly even when λ reaches 20. Arguably, a phase transition could still be reached for some even larger λ. However, as $\lambda \to \infty$ the infection rate β converges to $\beta' = \tau\langle a \rangle w$ (5.8). For the given parameters, it yields a basic reproduction number $R_0^{'} = \beta'/\gamma = 0.06895/0.07 = 0.985 < 1$. This means that a (relatively small) protective level of $w = 0.65$ will manage to impede long-lasting epidemics *for any walk rate* λ, in sharp contrast with the scenario where $w = 1$.

5.7.1 Dense Regimes

Even though the present work focuses on sparse scenarios, the spreading profile under dense regimes and the extent to which such dynamics can be reliably captured by the model arise as natural followup questions. Whereas the first strongly depends on numerical simulations (to be presented in this section), the latter brings

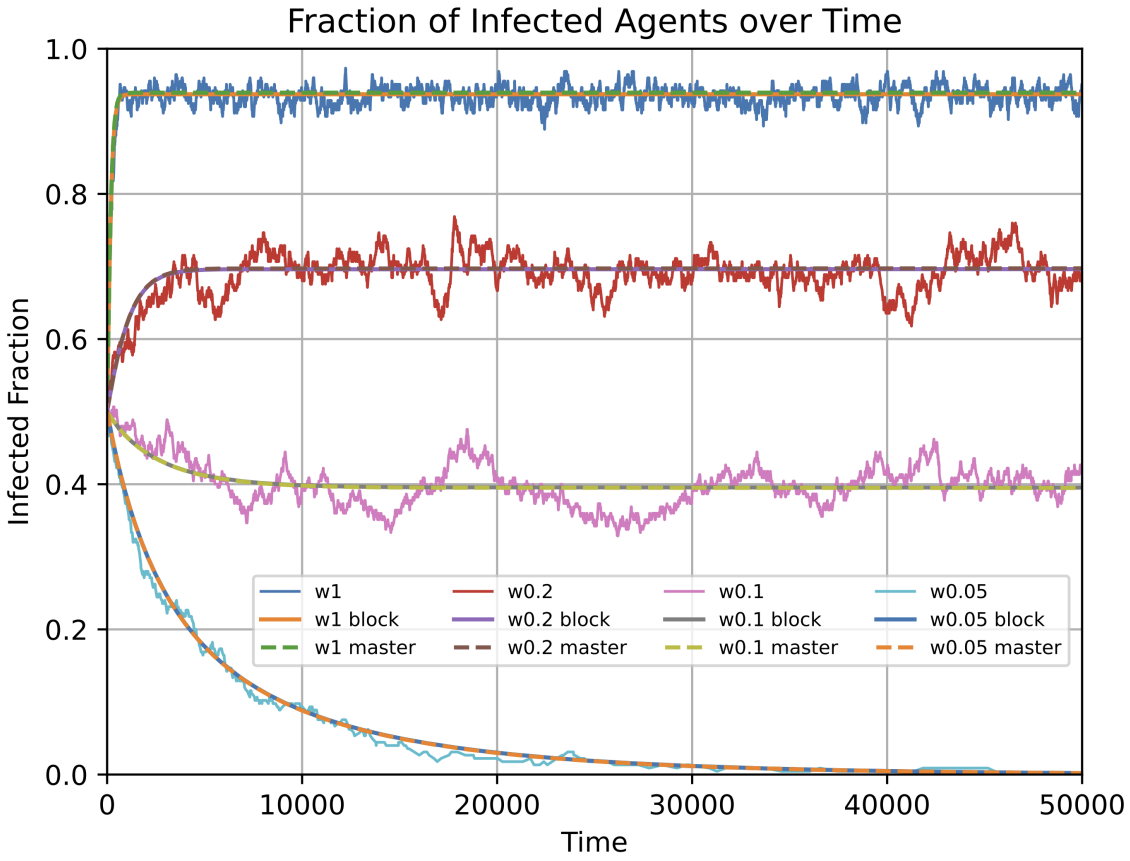


Figure 5.1: Fraction of infected agents over time for different risk-tolerance values (w) in a BA network, with their respective block-level (block) and network-level (master) predictors (n = 100K nodes, $\langle b \rangle \approx 21$, $K = 450$). For instance, the blue curve "w1" is the actual simulation of an epidemic when $w = 1$; the solid, orange curve "w1 block" is the prediction on the average behavior when $w = 1$ as given by numerically solving the system of equations in (4.14); likewise, the dashed, green curve "w1 master" expresses the numerical solution from the master equation (4.21) (which assumes MDC). Remarkably, both predictors are almost perfectly overlapped through all scenarios.

the infection probability (4.10) to a central spot in this discussion. As previously mentioned, (4.10) assumes sparsity, hence pairwise SI-contacts only. Although the task of deducing the infection probability for dense regimes lies beyond the scope of this work, an assessment on what is the gap between how these epidemics evolve and what the present model predicts when applied to them is of particular interest.

A first intuition is that a wrong prediction will almost surely be the case, but how exactly? It is clear that, prior to conducting experiments, this question must be divided into more concrete aspects in order for a proper understanding to be built. More specifically,

1. How different from each other would block-level (4.14) and network-level (4.21) predictors behave, regardless of the actual spreading?

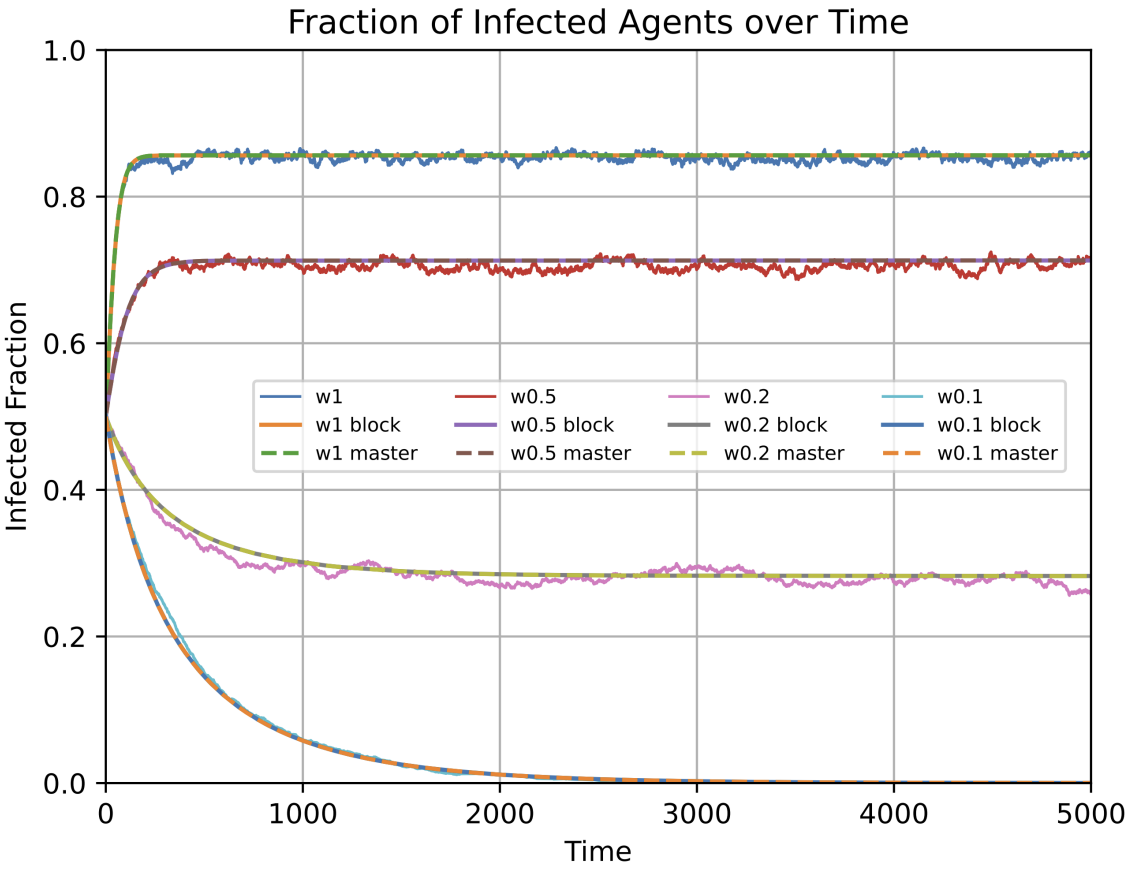


Figure 5.2: Fraction of infected agents over time for different risk-tolerance values (w) in a G(n,p) network, with their respective block-level and network-level predictors (n = 100K nodes, $\langle b \rangle \approx 21$, $K = 5000$). Here again, both predictors perfectly overlap each other through all scenarios.

2. How intensely do we expect these to under/overestimate the real spreading?

3. From a practical standpoint, how decisively will w influence the outcome of such epidemics?

Starting with the first question, such a difference will clearly depend on network structure. True, despite its provenience from MDC, (4.20) brought to evidence an otherwise hidden information on how heterogeneity plays a central role throughout the propagation. On the other hand, it also assumes state-level mixing, which is hardly the case for dense regimes. Both features considered, note that the network-level predictor (4.21) assumes that I-agents are better distributed across the network in comparison to block-level. (4.21) is thus the one expected to predict higher endemic levels. A more intuitive reasoning that leads to such a conclusion is to think of the *marginal virulence* every new incoming I-agent adds to the infective potential of an already hostile node: if a node hosts exactly one S- and one I-agent, then a new incoming I-agent offers 100% marginal virulence, as the risk of infecting the S-agent doubles. Likewise, a third incoming I-agent would impose a marginal

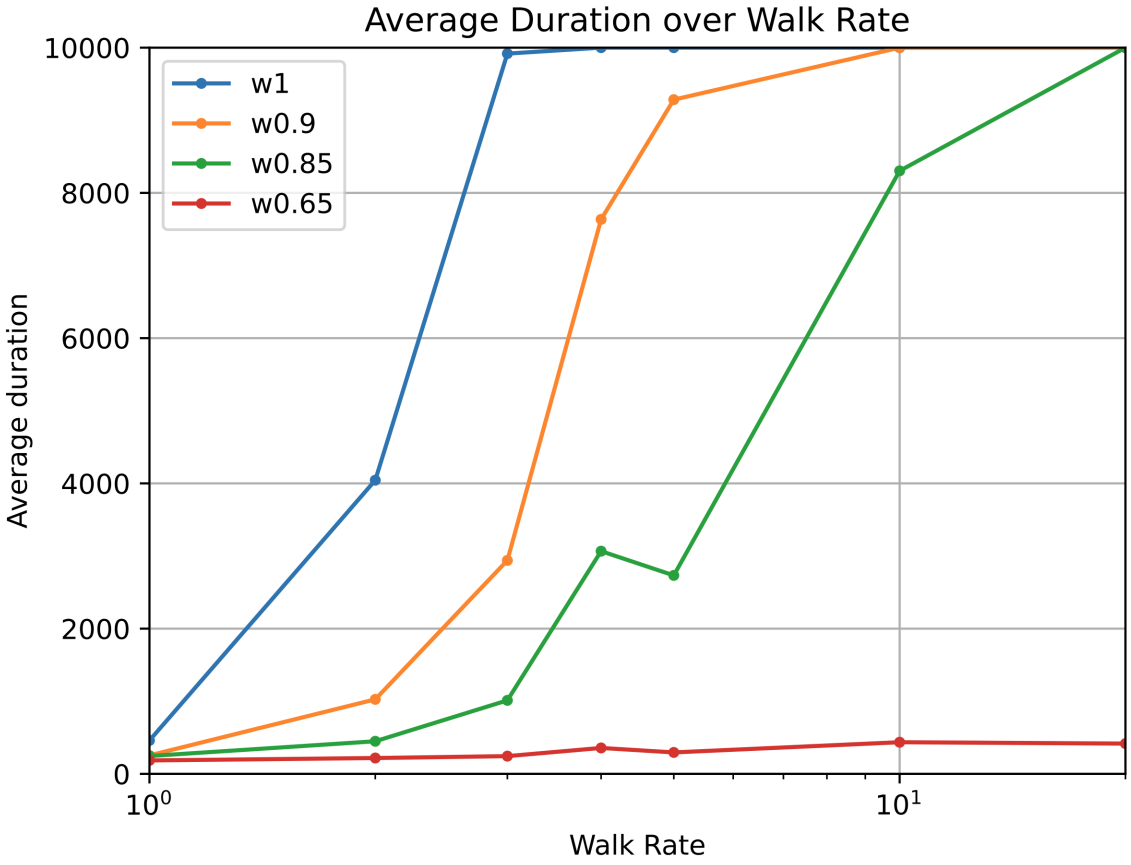


Figure 5.3: Epidemic's average duration over agent's walk rate (x-axis log-scaled). Each dot from each curve averages upon 10 runs for a same walk rate. Confidence intervals were omitted as the goal is not to directly compare curves, but to understand their average behavior in function of λ.

virulence of only 50% as the number of infectives now raises from 2 to 3, and so on. Clearly, a new incoming I-agent will at some point increase just negligibly the (already high) risk of infection, in a kind of saturation regime. This fact combined with heterogeneous scenarios where agents do *not* mix, mean that infectives are even more concentrated within the hubs in these cases. This, on the one hand, diminishes the risk of getting infected within those more remote nodes; on the other, every new infective coming into the hub contributes negligibly to its already strong infection force. This overall weaker spreading is better captured by the block-level predictor.

Moving forward to the second question, this is harder to conjecture when $w = 1$. Estimates on R_0, given by the relation between the force of infection β and the recovery rate γ are useful to some extent here. Besides, numerical results over different scenarios can be particularly revealing in this case, as we report shortly. Conversely, for very strong levels of PrE (e.g. $w = 0.001$), a short-lived epidemic will almost surely be the forecast, specially because such models assume sparsity.

By its turn, the third question—more than the first one—comes with a strong preliminary intuition: under enormously dense regimes, PrE is *not expected to no-*

ticeably change the course of an outbreak when I-agents are already made present in almost every node. In turn, it can also be thought of as a decisive preemptive artifact while the reach of an outbreak is still small. Numerical simulations (to be presented) can be profoundly enlightening here.

Now that a preliminary reasoning on what to expect from simulations over dense regimes has been formed, in what follows we proceed to report the simulations, and check its interesting results against the aforementioned expectations.

Figures to be presented in this section are for a different set of initial conditions. Instead of randomly infecting half the population, the set of early infectives *reside on a single node*. More specifically, we infect at time zero the entire population within the hub (i.e. the highest degree node) of a BA network, and then study the consequences of applying enormously different levels of PrE. All simulations to follow are run on the same BA network with $|V| = 10^4$ nodes and $K = 3 \times 10^5$ agents. Largest degree is 501 and average degree $\langle b \rangle \approx 21$.

For better clarity, results to follow are separated under two groups, summarized below:

1. *Pandemic regime.* This first group contains simulations at which the absence of protective efforts leads to massive spreading across a huge fraction of the population; we then apply a strong level of PrE to such a scenario and check whether it does meaningfully change the previously observed outcome.

2. *Otherwise endemic regime.* Scenarios that, even in the absence of PrE, give rise to epidemics that, although persistent, reach only a small portion of the agents. Here again, after running a simulation with $w = 1$ (i.e. no protection) we apply a strong PrE level to the system and observe how differently its related outbreak evolves, keeping in mind the expectations we have delineated previously in this section.

Pandemic regime

Figure 5.4 shows the fraction of infectives over time for a dense epidemic simulation with the aforementioned parameters. The figure contains 2 sets of 3 curves each. Each set expresses the simulation outcome for a given PrE level and its respective predictions as given by the master equation (4.21) and the block-level system (4.14). Levels of PrE adopted are $w = 1$ and $w = 0.001$.

Remarkably, PrE has a negligible influence on the overall fraction of infectives, *despite its huge intensity.* Not only is the hypercritical regime onset not delayed but both curves overlap both early and late in the epidemic. Indeed, note how the two epidemic curves ($w1$ and $w0.001$, respectively) end up converging to the same stationary level, despite a 1000× PrE factor separating both. This fact contradicts our

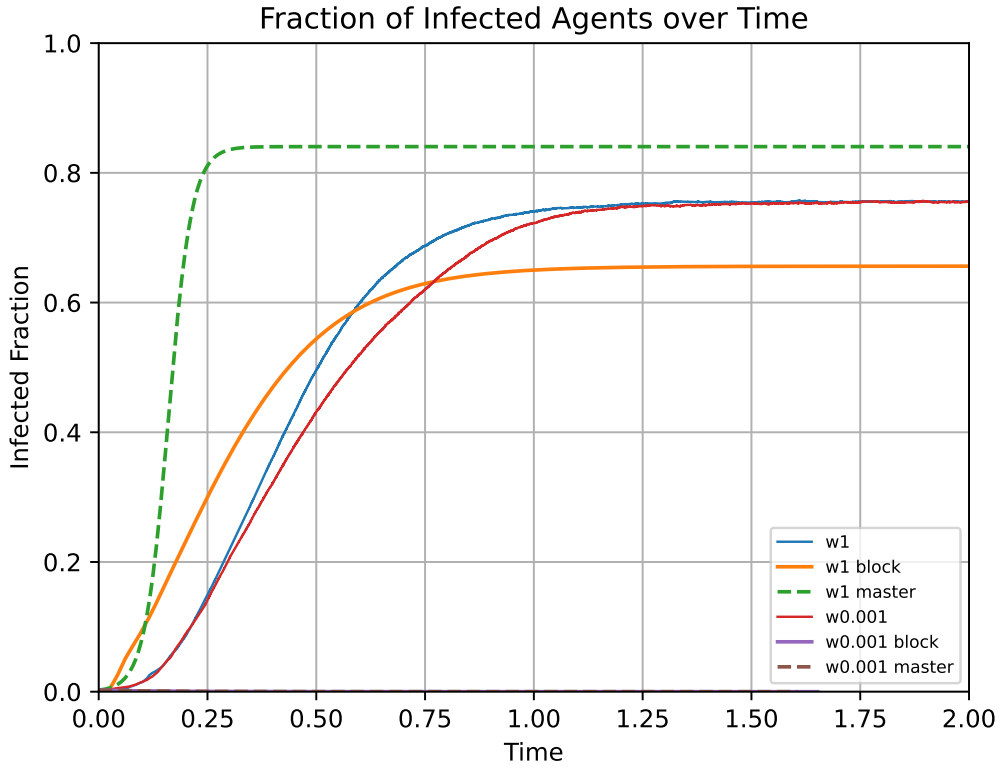


Figure 5.4: Fraction of infectives over time for an epidemic simulation on a dense regime in a BA network with $|V| = 10^4$ nodes, $K = 3 \times 10^5$ agents, largest degree = 501 and average degree $\langle b \rangle \approx 21$. For $w = 1$, $I_0 = 691$ (approx. only 0.23% of the agents) and for $w = 0.001$ the number of early spreaders is $I_0 = 707$, which is fairly close to the first case. The other parameters are $\lambda = 1, \tau = 1$, and $\gamma = 7$. Note that the stationary regime accounts for a pandemic, as the disease permeates a huge fraction of individuals.

previously formed intuition, as there are two separated facts here: first, if a massive number of infected agents are *already* widespread through the entire network, then S-agents will be always forced to hop into hostile nodes, as neither their current node nor their neighborhood are free from I-agents. This fact is actually easy to visualize and to conclude by intuition (as we did already). The counter-intuitive part is that the set of early spreaders here is quite small compared to K and PrE is very large, so that one would at least expect the epidemic onset to be noticeably delayed. This easy way through which the spreading made such profound inroads even in the presence of strong PrE is in agreement with results over static contact models, where nodes encode individuals rather than sites [48]. Therein, it is a well-known result that the epidemic threshold vanishes for highly heterogeneous networks.

Another major difference one may notice when comparing the present scenario with sparse regimes is a much smoother epidemic curve for the actual spreading (curves $w1$ and $w0.001$), indicating that dense scenarios are much less affected by

random fluctuations.

Interestingly, through all simulations for both cases we had both master and block-level predictors over/underestimating the actual spread, respectively, by nearly the same proportion when $w = 1$. This indicates that a reasonable prediction for dense scenarios without PrE could still be achieved up from the cumulative average of these two outcomes. On the other hand, when $w = 0.001$ both models make *grossly wrong predictions* of extremely short-lived outbreaks, *leading these curves to not even appear in the plot*, but this comes *as expected*, since both models assume sparsity.

Otherwise endemic regime

Contrariwise to the above results, where intuition and observance have converged to some extent (although not completely), Figure 5.5 comes with even more surprising outcomes.

Here, a preliminary intuition is: if endemic levels *without PrE* remain low (like, say, 30%) then, unless I-agents manage to span over all network nodes at any point in time (which is unlikely), a strong enough PrE must provoke safe sites to appear, thus gradually "absorbing" S-agents from neighboring regions, hence slowly but surely weakening the epidemic until it dies out.

Whereas a relatively constant, well behaved endemic level is shown to be the case for $w = 1$, a very peculiar spreading pattern emerges when $w = 0.001$. Interestingly, rich, oscillating dynamics arise as a result of the combination of strong PrE versus highly heterogeneous topology (with outbreak ignition at the hub). Indeed, note that when $w = 0.001$ only a small fraction of the population remains infected (around 2%), but in a curious oscillating fashion, seemingly narrowing up along with time. It turns out, however, that if we let the epidemic run for a longer time, then oscillations intensify back again from time to time. We demonstrate this behavior in Figure 5.6, which provides a magnified version of the *exact same curve* $w = 0.001$ from Figure 5.5 by restricting the y-axis to the interval $[0, 0.12]$ and also by extending the simulation time from 20 to 50.

This interesting phenomenon captured by Figure 5.6 offers new insights on why we frequently observe real-world seasonal epidemics that persist with only a small fraction of population: on the one hand, protective measures are taken (including mobility patterns); on the other hand, heterogeneous topology cause those more well connected nodes to gather sufficiently many agents to sustain local, long-lived spreading dynamics that from time to time reach out to neighboring regions.

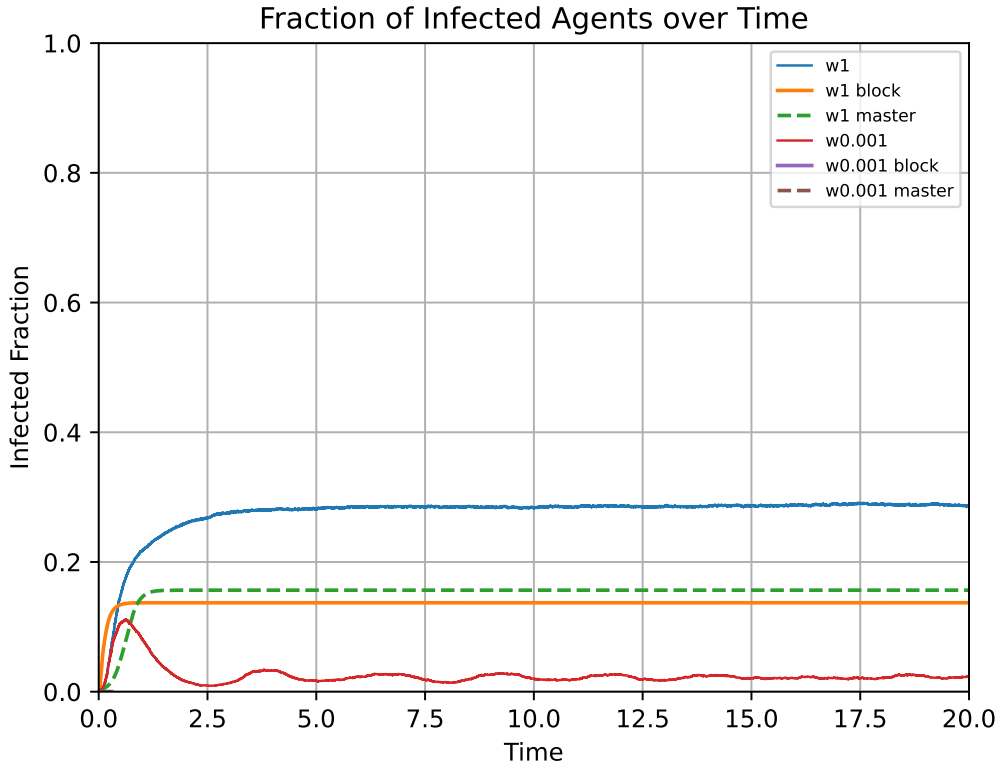


Figure 5.5: Fraction of infectives over time for an epidemic simulation on a dense regime in a BA network with $|V| = 10^4$ nodes, $K = 3 \times 10^5$ agents, largest degree $= 501$ and average degree $\langle b \rangle \approx 21$. Here again, $I_0 \approx 700$ for both curves. The other parameters are $\lambda = 1, \tau = 1$, and $\gamma = 37$.

5.8 Epidemic Simulator

A brief note on the epidemic simulator that conducted our series of experiments is worth it at this point. Our discrete-event simulator [65, 72] was entirely written in C++, with no inclusion of libraries for manipulating networks or for numerically solving ODE systems. Indeed, the author has made his own implementation of every aspect related to these features, and also has made the entire code publicly available.

Towards the key implementation-level decisions:

- Networks are represented and read as edge lists;
- By its turn, the numerical solver for the ODE systems implements the *4-th order Runge-Kutta* technique. In particular, the exact algorithm adopted is an extension of the 2-equation, first-order system [73], which is reproduced below:

4th-order Runge-Kutta method for 2-equation, first-order system:

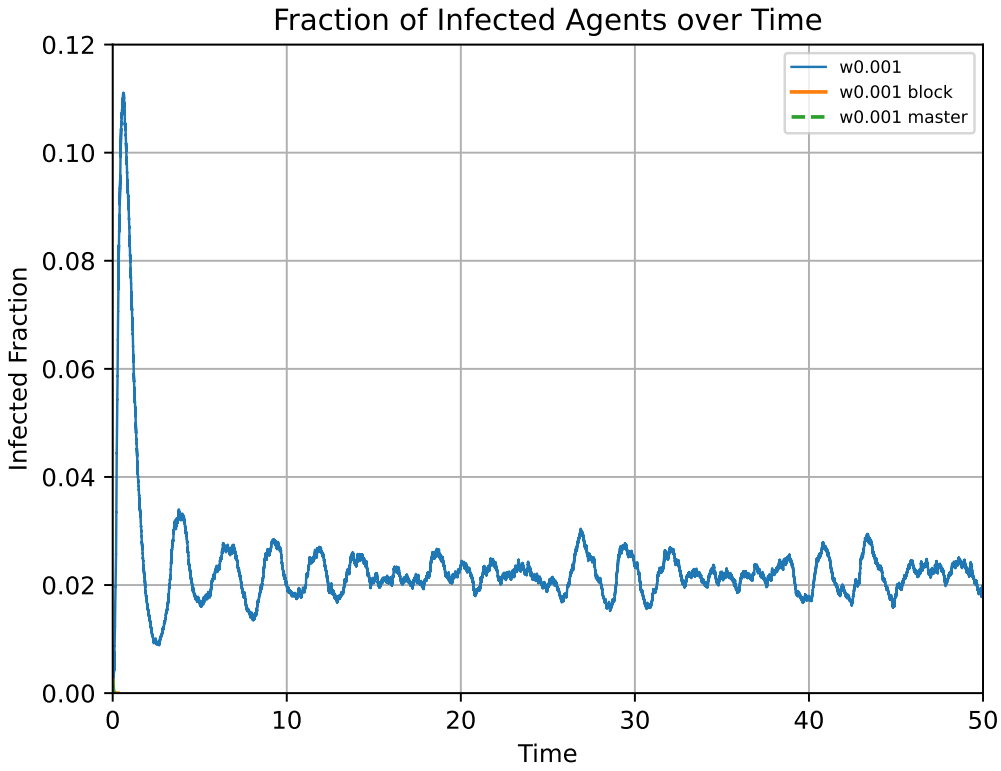


Figure 5.6: Magnified version of the curve $w = 0.001$ from Figure 5.5. The y-axis here is restricted to the interval $[0, 0.12]$. Also, simulation time is extended from 20 to 50. Extremely ephemeral propagation is forecast by both master and block-level predictors, hence their barely noticeable appearance (close to the origin).

For our (general) problem from class

$$\begin{aligned} dx/dt &= f(t,x,y), \quad x(t_0) = x_0 \\ dy/dt &= g(t,x,y), \quad y(t_0) = y_0 \end{aligned}$$

we get our approximate solution (x_n, y_n) at time t_n, $n = 1, 2, \ldots$ via the iteration of

$$\begin{aligned} x_{n+1} &= x_n + \tfrac{h}{6}(k_{n1} + 2k_{n2} + 2k_{n3} + k_{n4}) \\ y_{n+1} &= y_n + \tfrac{h}{6}(\ell_{n1} + 2\ell_{n2} + 2\ell_{n3} + \ell_{n4}), \end{aligned}$$

where the formulas for each of the k's and ℓ's are

$$\begin{aligned} k_{n1} &= f(t_n, x_n, y_n) \\ \ell_{n1} &= g(t_n, x_n, y_n) \\ k_{n2} &= f(t_n + h/2, x_n + \frac{1}{2}hk_{n1}, y_n + \frac{1}{2}h\ell_{n1}) \\ \ell_{n2} &= g(t_n + h/2, x_n + \frac{1}{2}hk_{n1}, y_n + \frac{1}{2}h\ell_{n1}) \\ k_{n3} &= f(t_n + h/2, x_n + \frac{1}{2}hk_{n2}, y_n + \frac{1}{2}h\ell_{n2}) \\ \ell_{n3} &= g(t_n + h/2, x_n + \frac{1}{2}hk_{n2}, y_n + \frac{1}{2}h\ell_{n2}) \\ k_{n4} &= f(t_n + h, x_n + hk_{n3}, y_n + h\ell_{n3}) \\ \ell_{n4} &= g(t_n + h, x_n + hk_{n3}, y_n + h\ell_{n3}). \end{aligned}$$

Chapter 6

Conclusion

This work proposed a simple agent-based model for continuous-time SIS epidemics on either complete graphs and also non-regular, degree-uncorrelated networks. Protective behavior came through biasing walkers towards safe sites, in state-dependent fashion: susceptible (S) agents avoid locations hosting infected (I) ones and vice-versa, at parameter-defined levels. Propagation was modeled as occurring through direct contact between S- and I- agents at a given node and depends on the total exposition time.

Through degree-block approximation, we provided an accurate, ODE-based predictor that explicitly embodies structural and protective information into a system typically orders of magnitude smaller than the network size; We have also provided the first theoretical evidence of a key structural aspect hitherto only observed empirically in agent-based models: larger *degree heterogeneity* boosts the spreading;

Remarkably, we also have found a special regime under which epidemics on networks with *arbitrary degree distribution* can be accurately predicted in $\Theta(1)$, not by a system but by *one closed-form, fast-computing equation*, namely the Matching Density Condition obtained via state-level mixing.

We have provided protective thresholds for disease-free steady states under either fixed and asymptotic rates for both walk and transmissibility, with a rigorous theoretical analysis.

Interestingly, we have found a law through which a simple random walk (SRW) can be conveniently combined with the degree function $d(\cdot)$ in order to converge in expected value to *any desired moment of a network's degree distribution.* Among other consequences, we answered a long open question on which value an SRW does actually approach when overestimating a network's average degree, which is now known to be $\langle b^2 \rangle / \langle b \rangle$. This finding raises research perspectives even beyond the scope of epidemics. For instance, methods for learning network structure via random walks could leverage this fact in order to design estimators with enhanced trade-off between simplicity and effectiveness, backed-up by theoretical results.

Last, our contributions also include the design and implementation of a publicly available network epidemic simulator whose numerical results validate the theoretical analysis on the proposed model.

6.1 Future Work

The following topics denote promising research directions.

- Of interest is a deeper investigation on time-varying protection effects, for which we have conducted only a preliminary study. The intention is to understand how an epidemic unfolds through *gradually reactive* protection, that is, the one that is more or less intense according to the number of infectives. A mathematical model—validated with simulation data—that captures such dynamics is currently a gap. Emerging issues include "Given a network (topology) and other initial conditions, up from what level of protective effort will the outbreak fade away?" It is also known that time-varying forces of infection give rise to *bistability*, a scenario where not only the model parameters but also the i_0 size of the initially infected population becomes fundamental to the correct prediction of the epidemic reach. An immediate implication is *seasonality*: once the epidemic is weakened, individuals relax their protective efforts likewise, thus allowing those few infections still active to recur in a new outbreak. The characterizations of topology and reactivity that give rise to the phenomenon are subject of investigation;

- Mathematically characterize PrE in high density regimes. Indeed, the numerical results obtained in this thesis are promising: they uncovered an interesting phenomenon captured by Figure 5.6, offering a new insight on why do we frequently observe real-world seasonal epidemics that persist with only a small fraction of population.

- Graph sampling methods that leverage the law of the n-th moment convergence (Proposition 1). This could potentially reveal a much simpler way for learning a network's structure up from relatively inexpensive random walks.

6.2 Acknowledgments

This work was partially funded by grants from CNPq and CAPES.

References

[1] BARABÁSI, A.-L., PÓSFAI, M. *Network science*. Cambridge, Cambridge University Press, 2016. ISBN: 9781107076266 1107076269. Disponível em: <http://barabasi.com/networksciencebook/>.

[2] RONALD, R. "Stochastic persistence and stationary distribution in an SIS epidemic model with media coverage.—Part I". p. 204–230, London, 1916. Proc. R. Soc. Lond. doi: http://doi.org/10.1098/rspa.1916.0007.

[3] OGILVY, K. W., G., M. A., THOMAS, W. G. "A contribution to the mathematical theory of epidemics", *Proc. R. Soc. Lond. A*, v. 115, 1927.

[4] DE SOUZA, R., FIGUEIREDO, D., DE A. ROCHA, A., et al. "Efficient network seeding under variable node cost and limited budget for social networks", *Information Sciences*, v. 514, pp. 369 – 384, 2020. ISSN: 0020-0255. doi: https://doi.org/10.1016/j.ins.2019.11.029. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0020025519310758>.

[5] CHANG, S., PIERSON, E., KOH, P. W., et al. "Mobility network models of COVID-19 explain inequities and inform reopening", *Nature*, v. 589, n. 7840, pp. 82–87, Jan 2021. ISSN: 1476-4687. doi: 10.1038/s41586-020-2923-3. Disponível em: <https://doi.org/10.1038/s41586-020-2923-3>.

[6] ATCHISON, C., BOWMAN, L. R., VRINTEN, C., et al. "Early perceptions and behavioural responses during the COVID-19 pandemic: a cross-sectional survey of UK adults", *BMJ Open*, v. 11, n. 1, 2021. ISSN: 2044-6055. doi: 10.1136/bmjopen-2020-043577. Disponível em: <https://bmjopen.bmj.com/content/11/1/e043577>.

[7] ZHONG, B.-L., LUO, W., LI, H.-M., et al. "Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey", *International journal of biological sciences*, v. 16, n. 10, pp. 1745–1752,

Mar 2020. ISSN: 1449-2288. doi: 10.7150/ijbs.45221. Disponível em: <https://doi.org/10.7150/ijbs.45221>.

[8] QIU, J., SHEN, B., ZHAO, M., et al. "A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations", *General psychiatry*, v. 33, n. 2, pp. e100213–e100213, Mar 2020. ISSN: 2517-729X. doi: 10.1136/gpsych-2020-100213. Disponível em: <https://doi.org/10.1136/gpsych-2020-100213>.

[9] LAZARUS, J. V., RATZAN, S. C., PALAYEW, A., et al. "A global survey of potential acceptance of a COVID-19 vaccine", *Nature Medicine*, v. 27, n. 2, pp. 225–228, Feb 2021. ISSN: 1546-170X. doi: 10.1038/s41591-020-1124-9. Disponível em: <https://doi.org/10.1038/s41591-020-1124-9>.

[10] CAPASSO, V., SERIO, G. "A generalization of the Kermack-McKendrick deterministic epidemic model", *Mathematical Biosciences*, v. 42, n. 1, pp. 43 – 61, 1978. ISSN: 0025-5564. doi: https://doi.org/10.1016/0025-5564(78)90006-8. Disponível em: <http://www.sciencedirect.com/science/article/pii/0025556478900068>.

[11] ZHANG, Z., WANG, H., WANG, C., et al. "Modeling Epidemics Spreading on Social Contact Networks", *IEEE Transactions on Emerging Topics in Computing*, v. 3, n. 3, pp. 410–419, Sep. 2015. ISSN: 2376-4562. doi: 10.1109/TETC.2015.2398353.

[12] WANG, Z., ANDREWS, M. A., WU, Z.-X., et al. "Coupled disease–behavior dynamics on complex networks: A review", *Physics of Life Reviews*, v. 15, pp. 1 – 29, 2015. ISSN: 1571-0645. doi: https://doi.org/10.1016/j.plrev.2015.07.006. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1571064515001372>.

[13] AMANTE, A., BALMER, C. "Italy in coronavirus lockdown as deaths soar and economy fades". 03 2020. Accessed: 2020-03-11.

[14] CAUCHEMEZ, S., BHATTARAI, A., MARCHBANKS, T., et al. "Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza", *PNAS*, v. 108, pp. 2825–30, 02 2011. doi: 10.1073/pnas.1008895108.

[15] DRAIEF, M., GANESH, A. "A Random Walk Model for Infection on Graphs: Spread of Epidemics & Rumours with Mobile Agents", *Discrete Event Dynamic Systems*, v. 21, n. 1, pp. 41–61, mar. 2011. ISSN: 0924-6703.

doi: 10.1007/s10626-010-0092-5. Disponível em: <https://doi.org/10.1007/s10626-010-0092-5>.

[16] LAM, H., LIU, Z., MITZENMACHER, M., et al. "Information Dissemination via Random Walks in D-Dimensional Space". In: *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '12, p. 1612–1622, USA, 2012. Society for Industrial and Applied Mathematics.

[17] IBRAHIM, P. S. Y. *Modelagem e Análise de Epidemias SIS em Redes baseadas em Passeios Aleatórios*. Tese de Mestrado, Federal University of Rio de Janeiro, Brazil, 2012.

[18] TAVARES, J. V. B. *Simulação Escalável de Epidemias SIS Baseadas em Passeios Aleatórios com Caracterização de Transições de Fase*. Tese de Mestrado, Federal University of Rio de Janeiro, Brazil, 2018.

[19] KEMPE, D., KLEINBERG, J., TARDOS, E. "Maximizing the Spread of Influence Through a Social Network". In: *ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, pp. 137–146, 2003. ISBN: 1-58113-737-0.

[20] ARTHUR, D., MOTWANI, R., SHARMA, A., et al. "Pricing Strategies for Viral Marketing on Social Networks". In: *Internet and Network Economics*, v. 5929, *Lecture Notes in Computer Science*, pp. 101–112, 2009. ISBN: 978-3-642-10840-2.

[21] KITSAK, M., GALLOS, L. K., HAVLIN, S., et al. "Identification of influential spreaders in complex networks", *Nature Physics*, v. 6, pp. 888–893, 2010.

[22] CHEN, W., WANG, Y., YANG, S. "Efficient influence maximization in social networks". In: *ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, pp. 199–208, 2009.

[23] HINZ, O., SKIERA, B., BARROT, C., et al. "Seeding Strategies for Viral Marketing: An Empirical Comparison", *Journal of Marketing*, v. 75, pp. 55–71, nov. 2011.

[24] ARAL, S., MUCHNIK, L., SUNDARARAJAN, A. "Engineering social contagions: Optimal network seeding in the presence of homophily", *Network Science*, v. 1, pp. 125–153, 2013. ISSN: 2050-1250.

[25] CHEN, D.-B., XIAO, R., ZENG, A., et al. "Path diversity improves the identification of influential spreaders", *Europhysics letters*, v. 104, jan. 2014. ISSN: 0295-5075.

[26] WANG, S., WANG, F., CHEN, Y., et al. "Exploiting social circle broadness for influential spreaders identification in social networks", *World Wide Web*, v. 18, n. 3, pp. 681–705, 2015.

[27] LIU, Y., WEI, B., DU, Y., et al. "Identifying influential spreaders by weight degree centrality in complex networks", *Chaos, Solitons and Fractals*, v. 86, 05 2016. ISSN: 0960-0779.

[28] BAE, J., KIM, S. "Identifying and ranking influential spreaders in complex networks by neighborhood coreness", *Physica A: Statistical Mechanics and its Applications*, v. 395, fev. 2014. ISSN: 0378-4371.

[29] LI, X., SMITH, J. D., DINH, T. N., et al. "TipTop: (Almost) Exact Solutions for Influence Maximization in Billion-Scale Networks", *IEEE/ACM Trans. Netw.*, v. 27, n. 2, pp. 649–661, 2019.

[30] LI, Y., FAN, J., WANG, Y., et al. "Influence Maximization on Social Graphs: A Survey", *IEEE Trans. Knowl. Data Eng.*, v. 30, n. 10, pp. 1852–1872, 2018.

[31] GONG, M., YAN, J., SHEN, B., et al. "Influence maximization in social networks based on discrete particle swarm optimization", *Inf. Sci.*, v. 367-368, pp. 600–614, 2016.

[32] SEIDMAN, S. B. "Network structure and minimum degree", *Social Networks*, v. 5, n. 3, set. 1983. ISSN: 0378-8733.

[33] UGANDER, J., BACKSTROM, L., MARLOW, C., et al. "Structural diversity in social contagion", *Proceedings of the National Academy of Sciences*, v. 109, n. 16, pp. 5962–5966, 2012.

[34] XU, W., LIANG, W., LIN, X., et al. "Finding top-k influential users in social networks under the structural diversity model", *Inf. Sci.*, v. 355-356, pp. 110–126, 2016.

[35] LESKOVEC, J., KRAUSE, A., GUESTRIN, C., et al. "Cost-effective outbreak detection in networks". In: *ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, pp. 420–429, 2007.

[36] MIYAUCHI, A., IWAMASA, Y., FUKUNAGA, T., et al. "Threshold Influence Model for Allocating Advertising Budgets". In: *International Conference on Machine Learning (ICML)*, pp. 1395–1404, 2015.

[37] NGUYEN, H., ZHENG, R. "On Budgeted Influence Maximization in Social Networks", *IEEE Journal on Selected Areas in Communications*, v. 31, n. 6, pp. 1084–1094, 2013.

[38] HAN, S., ZHUANG, F., HE, Q., et al. "Balanced Seed Selection for Budgeted Influence Maximization in Social Networks". In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 65–77, 2014.

[39] NGUYEN, H. T., THAI, M. T., DINH, T. N. "A Billion-Scale Approximation Algorithm for Maximizing Benefit in Viral Marketing", *IEEE/ACM Trans. Netw.*, v. 25, n. 4, pp. 2419–2429, 2017.

[40] DE SOUZA, R. C., FIGUEIREDO, D. R., DE A. ROCHA, A. A., et al. "Evaluation of Epidemic Seeding Strategies under Variable Node Costs". In: *SBC Workshop em Desempenho de Sistemas Computacionais e de Comunicação*, WPerformance '14, 2014.

[41] TANG, Y., XIAO, X., SHI, Y. "Influence maximization: near-optimal time complexity meets practical efficiency". In: *SIGMOD Conference*, pp. 75–86. ACM, 2014.

[42] TANG, Y., SHI, Y., XIAO, X. "Influence Maximization in Near-Linear Time: A Martingale Approach". In: *SIGMOD Conference*, pp. 1539–1554. ACM, 2015.

[43] PERLBERG, S. "Facebook Signs Deals With Media Companies, Celebrities for Facebook Live". jun. 2016. The Wall Street Journal.

[44] KORNOWSKI, L. "Celebrity Sponsored Tweets: What The Stars Get Paid For Advertising In 140 Characters". maio 2013. The Huffington Post.

[45] BROWN, K. "Here's How Much Celebrities Make in the Instagram Product Placement Machine". jan. 2016. Jezebel.

[46] NGUYEN, H. T., NGUYEN, T. P., VU, T. N., et al. "Outward Influence and Cascade Size Estimation in Billion-scale Networks". In: *SIGMETRICS (Abstracts)*, p. 63. ACM, 2017.

[47] NEWMAN, M. *Networks: An Introduction*. USA, Oxford University Press, Inc., 2010. ISBN: 0199206651.

[48] PASTOR-SATORRAS, R., VESPIGNANI, A. "Epidemic Spreading in Scale-Free Networks", *Physical review letters*, v. 86, pp. 3200–3, 05 2001. doi: 10.1103/PhysRevLett.86.3200.

[49] NEWMAN, M. E. J. *Networks: An Introduction*. USA, Oxford University Press, 2010. ISBN: 0199206651.

[50] FUNK, S., SALATHÉ, M., JANSEN, V. "Modelling the Influence of Human Behaviour on the Spread of Infectious Diseases: A Review", *Journal of the Royal Society, Interface / the Royal Society*, v. 7, pp. 1247–56, 09 2010. doi: 10.1098/rsif.2010.0142.

[51] HYMAN, J. M., LI, J. "Behavior Changes in SIS STD Models with Selective Mixing", *SIAM Journal on Applied Mathematics*, v. 57, n. 4, pp. 1082–1094, 1997. doi: 10.1137/S0036139995294123. Disponível em: <https://doi.org/10.1137/S0036139995294123>.

[52] TCHUENCHE, J., DUBE, N., BHUNU, C., et al. "The impact of media coverage on the transmission dynamics of human influenza", *BMC public health*, v. 11 Suppl 1, pp. S5, 02 2011. doi: 10.1186/1471-2458-11-S1-S5.

[53] CAI, Y., KANG, Y., BANERJEE, M., et al. "A stochastic SIRS epidemic model with infectious force under intervention strategies", *Journal of Differential Equations*, v. 259, n. 12, pp. 7463 – 7502, 2015. ISSN: 0022-0396. doi: https://doi.org/10.1016/j.jde.2015.08.024. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0022039615004271>.

[54] GUO, W., CAI, Y., ZHANG, Q., et al. "Stochastic persistence and stationary distribution in an SIS epidemic model with media coverage", *Physica A: Statistical Mechanics and its Applications*, v. 492, pp. 2220 – 2236, 2018. ISSN: 0378-4371. doi: https://doi.org/10.1016/j.physa.2017.11.137. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0378437117311767>.

[55] GRANELL, C., GÓMEZ, S., ARENAS, A. "Dynamical Interplay between Awareness and Epidemic Spreading in Multiplex Networks", *Phys. Rev. Lett.*, v. 111, pp. 128701, Sep 2013. doi: 10.1103/PhysRevLett.111.128701. Disponível em: <https://link.aps.org/doi/10.1103/PhysRevLett.111.128701>.

[56] MAO, L., YANG, Y. "Coupling infectious diseases, human preventive behavior, and networks – A conceptual framework for epidemic modeling", *Social Science & Medicine*, v. 74, n. 2, pp. 167 – 175, 2012.

ISSN: 0277-9536. doi: https://doi.org/10.1016/j.socscimed.2011.10.012. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0277953611006551>.

[57] VOLZ, E., MEYERS, L. "Susceptible-infected-recovered epidemics in dynamic contact networks", *Proceedings. Biological sciences / The Royal Society*, v. 274, pp. 2925–33, 09 2007. doi: https://doi.org/10.1098/rspb.2007.1159.

[58] EAMES, K. T. D., TILSTON, N. L., BROOKS-POLLOCK, E., et al. "Measured Dynamic Social Contact Patterns Explain the Spread of H1N1v Influenza", *PLOS Computational Biology*, v. 8, n. 3, pp. 1–8, 03 2012. doi: 10.1371/journal.pcbi.1002425. Disponível em: <https://doi.org/10.1371/journal.pcbi.1002425>.

[59] ROBINSON, S., EVERETT, M., CHRISTLEY, R. "Recent network evolution increases the potential for large epidemics in the British cattle population", *Journal of the Royal Society, Interface / the Royal Society*, v. 4, pp. 669–74, 09 2007. doi: 10.1098/rsif.2007.0214.

[60] ZHOU, J., XIAO, G., CHEONG, S. A., et al. "Epidemic reemergence in adaptive complex networks", *Phys. Rev. E*, v. 85, pp. 036107, Mar 2012. doi: 10.1103/PhysRevE.85.036107. Disponível em: <https://link.aps.org/doi/10.1103/PhysRevE.85.036107>.

[61] YANG, H.-X., TANG, M., WANG, Z. "Suppressing epidemic spreading by risk-averse migration in dynamical networks", *Physica A: Statistical Mechanics and its Applications*, v. 490, pp. 347 – 352, 2018. ISSN: 0378-4371. doi: https://doi.org/10.1016/j.physa.2017.08.067. Disponível em: <http://www.sciencedirect.com/science/article/pii/S037843711730804X>.

[62] LORIG, F., JOHANSSON, E., DAVIDSSON, P. "Agent-Based Social Simulation of the Covid-19 Pandemic: A Systematic Review", *Journal of Artificial Societies and Social Simulation*, v. 24, n. 3, pp. 5, 2021. ISSN: 1460-7425. doi: 10.18564/jasss.4601. Disponível em: <http://jasss.soc.surrey.ac.uk/24/3/5.html>.

[63] BONABEAU, E. "Agent-based modeling: Methods and techniques for simulating human systems", *Proceedings of the National Academy of Sciences*, v. 99, n. suppl_3, pp. 7280–7287, 2002. doi: 10.1073/pnas.082080899. Disponível em: <https://www.pnas.org/doi/abs/10.1073/pnas.082080899>.

[64] BRITTON, T., HOUSE, T., LLOYD, A., et al. "Five challenges for stochastic epidemic models involving global transmission", *Epidemics*, v. 10, pp. 54–7, 2015. doi: 10.1016/j.epidem.2014.05.002.

[65] ROSS, S. "Chapter 7 - The Discrete Event Simulation Approach". In: *Simulation*, fifth edition ed., Academic Press, pp. 111 – 134, 2013. ISBN: 978-0-12-415825-2. doi: https://doi.org/10.1016/B978-0-12-415825-2.00007-3. Disponível em: <http://www.sciencedirect.com/science/article/pii/B9780124158252000073>.

[66] LESKOVEC, J., KREVL, A. "SNAP Datasets: Stanford Large Network Dataset Collection". http://snap.stanford.edu/data, 2014.

[67] DA F. COSTA, L., TRAVIESO, G. "Exploring Complex Networks through Random Walks", *Physical review. E, Statistical, nonlinear, and soft matter physics*, v. 75, pp. 016102, 02 2007. doi: 10.1103/PhysRevE.75.016102.

[68] BARRAT, A., BARTHÉLEMY, M., VESPIGNANI, A. *Dynamical Processes on Complex Networks*. Cambridge, Cambridge University Press, 2008. doi: 10.1017/CBO9780511791383.

[69] MOLLOY, M., REED, B. "A critical point for random graphs with a given degree sequence", *Random Structures & Algorithms*, v. 6, n. 2-3, pp. 161–180, 1995. doi: https://doi.org/10.1002/rsa.3240060204. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.3240060204>.

[70] CASTELLANO, C., PASTOR-SATORRAS, R. "Thresholds for Epidemic Spreading in Networks", *Phys. Rev. Lett.*, v. 105, pp. 218701, Nov 2010. doi: 10.1103/PhysRevLett.105.218701. Disponível em: <https://link.aps.org/doi/10.1103/PhysRevLett.105.218701>.

[71] LEVIN, D. A., PERES, Y., WILMER, E. L. *Markov chains and mixing times*. American Mathematical Society, 2006.

[72] DE SOUZA, R. C. "Simulator on Network Epidemics with Mobile Agents and PrE". Dec 2023. Disponível em: <https://github.com/rchiesse/randomWalk/tree/multiSchedule>.

[73] SCOFIELD, T. "Runge-Kutta Notes". Disponível em: <https://sites.calvin.edu/scofield/courses/m231/materials/rungeKuttaFormulas.pdf>.

Appendix A

Infection Probability

The infection probability as stated in Equation 4.10 comes from the notion of "competing exponentials", as follows. Without prejudice, let us assume no self-loops, thus $\lambda_b = \lambda$ i.e., the exit rate matches the walk rate. An infection then occurs when the transmission event fires (at rate τ) prior to a walk event (at rate 2λ since it may come from both agents, at rate λ each). Moreover, 2λ and τ relate to *independent* events. Let us notate $A =$ "infection" and $B =$ "walk away". Since both are independent exponentials, their joint density is

$$p(a, b) = \tau 2\lambda e^{-(\tau a + 2\lambda b)}, \tag{A.1}$$

from which we want $P(A < B)$. Thus,

$$\begin{aligned} P(A < B) &= \int_0^\infty \int_0^b p(a, b) da db \\ &= \int_0^\infty 2\lambda e^{-2\lambda b} \int_0^b \tau e^{-\tau a} da db \\ &= \int_0^\infty 2\lambda e^{-2\lambda b}(1 - e^{-\tau b}) db \\ &= \int_0^\infty 2\lambda e^{-2\lambda b} db - \int_0^\infty 2\lambda e^{-(\tau + 2\lambda) b} db \\ &= 1 - \frac{2\lambda}{2\lambda + \tau} \\ &= \frac{\tau}{2\lambda + \tau}. \end{aligned} \tag{A.2}$$